Universidade Federal do Rio de Janeiro
Centro de Tecnologia - Escola Politécnica
Departamento de Engenharia Eletrônica e de Computação

Disciplina: Telefonia 2003/2
Professor: Ricardo Rhomberg Martins

Apostila de Telefonia

Autores

Alexandre A. Barcellos Rodrigues
Alvaro M. Machado Neto
Aurélio C. de Magalhães
Carlos Fernando Teodósio Soares
Carlos Magno R. Vasques
Celso Alexandre S. de Alvear
Daniel Francisco Pinto
Daniel Levitan
Eduardo Rapoport
Fábio Carvalho dos Santos
Fábio Soares Gomes
Fernando J. Figueiredo Nazareth
Fernando Wilson Lugon
Flávia Almada Horta Marques
Gabriel Barros Durante
Gabriel Epjztejn
Guilherme Mello de Moura
Guilherme Pastor Garnier
Gustavo Luis A. de Carvalho
Isidoro C. Raposo de Almeida Filho
Igor Monteiro Moraes
José Antonio Casemiro Neto
Julio Cesar Bueno Filho
Leandro C. da Silva Annibal
Leonardo Duzzi de Lima
Lucas Lopes Alenquer
Marcela Ribeiro G. da Trindade
Marcelo A. Schmalter Soares
Marina Cruz
Renato Feiner
Ricardo dos Santos Souza

Roberto Barreto de Moraes

Rodrigo Franceski Prestes

Rômulo Mendes Cardoso

Ronald Pereira Mascarenhas

Silvia Galvão Lyra

Tiago Villela Hosken

Thiago Morra R. Fonseca

Organizador

Roberto Barreto de Moraes.

Orientador

Ricardo Rhomberg Martins, D. Sc.

Apresentação

O presente trabalho foi conseqüência de uma das avaliações do curso de Telefonia do Depto de Engenharia Eletrônica e de Computação da Escola Politécnica da UFRJ no 2º. Semestre de 2003. No primeiro dia de aula apresentei aos alunos o livro “Telefonia Digital” de Marcelo Sampaio de Alencar, fiz algumas observações e expressei meu pesar por não ter encontrado nenhum texto apropriado para quem tinha feito todas as cadeiras anteriores a esta no nosso curso.

Era a 1ª. vez que a cadeira seria oferecida depois de muitos anos no Departamento. Durante este período o curso passou por uma profunda reforma curricular que inclusive mudou seu nome. O “coração” do curso atual são as cadeiras de Circuitos Lógicos, Técnicas Digitais, Arquitetura de Computadores, Microprocessadores, etc... Deste ponto de vista parecia-me que uma boa proposta para um curso de Telefonia seria a criação de um “hardware” que possibilitasse o uso de um PC como central telefônica. Uma prova de que os alunos aproveitaram o que tinham visto nas cadeiras citadas foi que eles não aceitaram que o projeto fosse parte da avaliação. No entanto “encararam” a missão de preparar um texto que servisse de base para cursos futuros.

Alguns capítulos acabaram ficando parecidos com o livro do Alencar. Outros foram originais mas não atingiram o nível que deve ter um curso de Engenharia. Da minha parte destaco o capítulo de redes de computadores que, sem fazer as referências que seriam desejáveis num texto de Telefonia, ficou muito bom. O capítulo sobre Telefonia Celular e o apêndice A me pareceram perfeitos e o apêndice B, apesar de muito bom, eu acho que deveria ser exigido como pré-requisito para o curso (seu conteúdo é visto nas cadeiras de Comunicações Analógicas e Comunicações Digitais que o antecedem).

“Apesar dos pesares” estão todos de parabéns. O resultado final ficou bom e espero que as turmas dos próximos semestres consigam fazer algo semelhante. E o hardware que vocês acabaram fazendo sem valer nota tornou-se um distintivo do nosso curso em relação aos de outras Faculdades.

Ricardo Rhomberg Martins. Professor

Sumário

Noções de Acústica e Telefonia

1.1. INTRODUÇÃO – UMA BREVE HISTÓRIA DA TELEFONIA

1.1.1. Eletricidade e Magnetismo

A telefonia, assim como inúmeras outras áreas da Engenharia Elétrica e ciência modernas, teve seu início com a descoberta e desenvolvimento da Eletricidade e do Magnetismo, cujos estudos foram iniciados em 1830 pelo inglês Michael Faraday.

Como nos mais variados temas de pesquisa, a pergunta da época (e realmente satisfatória) era: "Para que serve Eletricidade e Magnetismo?". De 1830 a 1900 os cientistas foram descobrindo as aplicações deste ramo da Ciência, tornando-as indispensáveis no dia-a-dia. O Telefone, inventado em 1876, foi uma das primeiras aplicações.

A primeira invenção foi feita em 1837 por Samuel Morse: o Telégrafo Elétrico, que permite a comunicação entre dois pontos ligados por uma fiação condutora. Daí surgiram duas tecnologias de Telecomunicações: uma que suprimia a fiação condutora, usando transmissões de ondas eletromagnéticas no ar (Rádio, Televisão), e outra que mantinha a fiação (Telégrafo, Telefone, Fax). Com o Telefone Móvel, sem fiação, e a Televisão a cabo, estas tecnologias voltaram a se encontrar.

O começo da Telefonia foi confuso, sem uma noção clara do que poderia ser transmitido. Assim, logo se tentou a transmissão de cópias de documentos (Fax, em 1843 por Alexandre Bain), que foi conseguida, mas teve de esperar o desenvolvimento da Eletrônica para se tornar popular. O alemão Philipp Reis obteve em 1860 resultados razoáveis com um aparelho que transmitia música através de fiação, mas por ser exclusivamente para música teve divulgação limitada.

Um ponto muito importante na história das telecomunicações se deve a Alexander Graham Bell. Um pouco de sua história torna interessante nossa leitura. Passemos a ela.

1.1.2. A história de Bell

Alexander Graham Bell nasceu em Edimburgo, na Escócia, em 1847. Seu interesse na reprodução de sons vocais pode ser atribuído em parte ao trabalho de seu pai, Alexander Melville Bell, que foi um especialista na correção da fala e no ensino de deficientes auditivos.

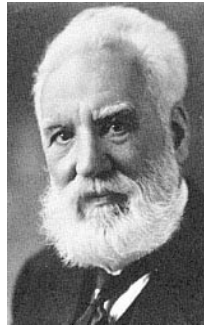



Figura 1-1.1: Alexander Graham Bell

Em 1871, Bell imigrou para a cidade norte-americana de Boston, estado de Massachusetts. Foi professor de fisiologia vocal na Universidade de Boston, onde, em 1873, iniciou convenções para professores de surdos. Foi também o fundador da Associação Americana para Promoção do Ensino da Fala aos Deficientes Auditivos.

De 1873 a 1876, Bell envolveu-se em diversos experimentos que posteriormente levaram-no a descobrir o telefone. Sua pesquisa foi financiada pelos pais de dois de seus alunos; um deles, chamado Gardiner Hubbard, tinha uma filha surda, Mabel, que posteriormente tornou-se esposa de Bell.

A concepção de Bell sobre a teoria do telefone era a seguinte: a intensidade de uma corrente elétrica pode ser modificada em função da variação da densidade do ar durante a produção do som. Ao contrário do telégrafo que usa uma corrente intermitente, o telefone exige uma corrente contínua com intensidade variável. Em 1874, Bell inventou um telégrafo harmônico para transmitir diversas mensagens simultaneamente por um único fio, e um receptor telefônico-telegráfico. Na tentativa de reproduzir a voz humana eletricamente, ele tornou-se um especialista na transmissão elétrica de ondas. Enquanto Bell teve a idéia, seu parceiro, Thomas Watson, produziu e montou o equipamento.

Elisha Gray e Alexandre Graham Bell descobrem que estão trabalhando no mesmo projeto: a invenção do telefone. Enquanto Bell buscava solução pelo lado acústico, Gray analisava as possibilidades de uso da corrente elétrica.

Ainda no ano de 1874 Gray concluiu seu projeto mais extraordinário: com um diafragma de aço diante de um eletroímã, havia construído um receptor praticamente igual aos dos telefones atuais. Mas não dispunha de transmissor. Em 14 de fevereiro de 1876, por extrema coincidência, os dois ingressam com o pedido de patente. Com apenas uma vantagem para Bell: seu requerimento chegara a "U.S. Patent Office" duas horas antes de Gray.

Até aqui ninguém havia conseguido realmente conversar pelo telefone. Gray parece perder as esperanças e considera-se derrotado pela diferença de duas horas na primazia do registro. Embora tivesse excelentes idéias e estivesse no caminho certo, nada faz para desenvolver os projetos que elaborou. E perde, assim, a corrida nos anos críticos que se seguem.

A patente lhe foi concedida em 7 de março. Bell continuou com seus experimentos para melhorar a qualidade do aparelho. Acidentalmente, as primeiras palavras ditas através de um telefone foram de Bell: "Sr. Watson, venha aqui. Eu preciso de sua ajuda" em 10 de março de 1876.

Pouco após receber a patente, Bell apresentou sua invenção na Exposição do Centenário, na Filadélfia. Seu aparelho gerou grande interesse público e recebeu um prêmio no evento. O Imperador Dom Pedro II esteve presente e fez uma encomenda de 100 aparelhos para o Brasil.

Demonstrações consecutivas superaram o ceticismo público sobre o telefone. A primeira conversa telefônica externa foi entre Bell e Watson, em 9 de outubro de 1876, e ocorreu entre as cidades de Boston e Cambridge, em Massachusetts. Em 1877, foi instalado o primeiro telefone residencial e foi conduzida uma conversa entre Boston e Nova Iorque, usando linhas de telégrafo. No mesmo ano, enquanto estava em lua-de-mel, Bell introduziu o telefone na Inglaterra e na França.

1.1.3. A telefonia no Brasil

O Brasil figura entre os primeiros países do mundo a ter, em seu território, telefones em funcionamento. A princípio, o aparelho circulava, provavelmente, mais como curiosidade científica do que com o caráter prático de hoje.

Existe alguma polêmica em torno de qual seria o primeiro telefone a chegar ao país. Uma versão dá conta de que teria sido instalado em 1877 (um ano depois de sua apresentação

na Exposição de Filadélfia), e funcionava na casa comercial "O Grande Mágico", no Beco do Desvio, depois rua do Ouvidor nº 86, ligando a loja ao quartel do Corpo de Bombeiros (Rio).

Outra, afirma que D. Pedro II teria recebido o primeiro aparelho como presente do próprio Graham Bell e teria começado a funcionar em janeiro de 1877, no Palácio de São Cristóvão (hoje Museu Nacional), na Quinta da Boa Vista. O aparelho utilizava uma linha até o centro da cidade e fora construído nas oficinas da "Western and Brazilian Telegraph Company".

Figura 1-2.2: D. Pedro na Exposição da Filadélfia

O número de aparelhos aumentava a cada ano, sem que fossem tomadas providências para que os serviços funcionassem em larga escala. Só em 15 de novembro de 1879 é que um decreto Imperial outorga a primeira concessão para exploração dos serviços no Brasil, concedendo a Paul Mackie, que representava os interesses da "Bell Telephone Company", licença para construir e operar linhas telefônicas na capital do Império (Rio), e na cidade de Niterói. Nesse período as linhas não eram cobradas dos assinantes, que pagavam apenas uma taxa anual ou mensal para sua utilização. Desse ano até o final do Império, seguiram-se inúmeros decretos de regulamentação. Todos procuravam ordenar a prestação do serviço, compatibilizando-o com a infra-estrutura de telégrafos já existente e distribuindo concessões nas várias regiões.

Em 1892, Lars Magnus Ericsson, o sueco que em 1876 fundou a L. M. Ericsson, iniciando seus trabalhos numa oficina modesta de consertos e reformas de telégrafos, industrializa o primeiro aparelho telefônico em que o transmissor e receptor (bocal e auricular) estão acoplados numa única peça, criado por Anton Avéns e Leonard Lundqvist,

em 1884, dando origem ao monofone. São os chamados, no Brasil, de "pés de ferro", e na Argentina, de "telefone aranha".

No mesmo ano (1892), Almon Brown Strowger, empresário funerário de grande habilidade na construção de aparelhos elétricos e telegráficos, cria o embrião da primeira central telefônica automática.

Seu objetivo era simples e claro: livrar-se da concorrência desleal de uma telefonista de La Porte, Indiana, esposa de outro proprietário de empresa funerária, que não completava as ligações de possíveis clientes para seu estabelecimento; a telefonista se "equivocava" quando alguém pedia uma ligação para a funerária de Strowger. A primeira central automática do mundo tinha apenas 56 telefones.

Com a chegada da República em 1889, poucas alterações foram observadas na relação entre poder público e prestadores de serviço telefônico. A alteração mais significativa foi uma maior rigidez e controle do Estado com relação ao valor cobrado pelo serviço. Os preços foram estabelecidos em decreto de 26 de março de 1890. Não obstante esse maior controle, todos os contratos anteriormente celebrados pelo governo Imperial foram honrados à risca, demonstrando como Império e República mantiveram com estas empresas, a maioria de capital estrangeiro, uma relação muito parecida.

A automatização se fará gradativamente. Só nos primeiros anos do século XX, as principais cidades norte-americanas instalam suas centrais automáticas. Em 1913, Paris conta com 93 mil telefones manuais. Nova York, contudo, já dispõe de uma rede de 500 mil telefones, mas a automatização total só ocorrerá a partir de 1919.

No Brasil, a cidade de Porto Alegre é a primeira a inaugurar uma central automática em 1922 (a terceira das Américas, depois de Chicago e Nova York). A segunda do Brasil ainda será uma cidade gaúcha: Rio Grande, em 1925 - antes de Paris e Estocolmo. A estação pioneira da capital paulista foi inaugurada em julho de 1928, com o prefixo "5", na Rua Brigadeiro Galvão, na área do Centro "Palmeiras". Eram 9 mil terminais de fabricação norte americana "Automatic Electric", que funcionaram ininterruptamente, até meados de 1997.

O telefone tinha nos primeiros anos do século XX, pelo menos para grande parte da população, um interesse muito reduzido. O significado do telefone em termos de mudança, não tinha sido, ainda, captado pela sociedade. Tanto é verdade, que a maioria das empresas de telefonia sofria problemas para tornar seus negócios rentáveis. Era comum a companhia pedir a um cidadão para aceitar em sua casa, gratuitamente, a título de experiência, um aparelho telefônico. Também era comum que este fosse devolvido imediatamente após o menor "acidente" (uma chamada recebida tarde da noite, por exemplo).

No entanto, este não é um privilégio do telefone. Muitas invenções, que hoje nos parecem fundamentais, já sofreram com isso, mas como o século XX caracterizou-se pela capacidade de "criar" necessidades. Em poucos anos o telefone foi ganhando prestígio, difundido pelas várias regiões do país. Enfim, tornou-se uma necessidade, e hoje, todos sabemos, o telefone é, para muitas pessoas, realmente necessário e para instituições como bancos, empresas, etc., é um recurso sem o qual se tornaria simplesmente impossível operar seus serviços.

Isso é muito curioso no que se refere aos primeiros anos do telefone: a forma como vai se incorporando à vida dos homens, tornando-se parte do seu mundo, mesclando-se a suas atividades e tomando seu espaço, até tornar-se o serviço imprescindível que é hoje.

1.1.4. Rumos do Invento e de seu Inventor

Alexander Graham Bell ofereceu a venda de sua invenção para a companhia Western Union and Telegraph, por $100.000, mas a empresa recusou. E então, em julho de 1877, ele e seus parceiros fundaram sua própria companhia, que antecedeu a atual gigante global das telecomunicações, a Companhia Americana de Telefone e Telégrafo (American Telephone and Telegraph Company – AT&T). O telefone tornou-se um enorme sucesso econômico e a AT&T uma das maiores e melhores companhias mundiais.

Em 1879, Bell e sua esposa venderam aproximadamente 15% das ações de sua companhia telefônica. Aparentemente eles não perceberam os lucros fantásticos que a companhia iria gerar, porque sete meses depois eles já haviam vendido a maioria de suas ações ao preço médio de 250 dólares por ação. Em novembro, cada ação da empresa estava sendo vendida ao preço de $1000. Apesar de terem vendido sua participação a um preço historicamente baixo, em 1883, Bell e sua esposa já haviam lucrado por volta de 1 milhão de dólares. (Devemos lembrar que, naquela época, esta quantia representava muito mais do que vale hoje).

Apesar de ter se tornado rico com sua invenção, Bell nunca parou de conduzir pesquisas e foi capaz de inventar outros aparelhos úteis. Ele tinha interesses diversos em suas pesquisas, mas sua principal preocupação era a cura da surdez, provavelmente porque sua esposa sofria deste problema. O casal Graham teve dois filhos e duas filhas, mas, infelizmente, ambos os meninos morreram enquanto crianças.

As contribuições de Bell para o desenvolvimento da ciência são imensuráveis. A revista americana Science, que posteriormente tornou-se órgão oficial da Associação Americana para Avanço da Ciência, foi fundada em 1880 devido aos esforços do próprio Bell. Ele também foi presidente da Sociedade Geográfica Nacional de 1896 a 1904. A aviação foi um de seus primeiros interesses, e ele também realizou contribuições importantes para este campo da ciência.

Dentre todos os sucessos de sua vida, o telefone permanece como o principal. Poucas invenções humanas tiveram tanta influência no mundo. Além disso, Bell foi a primeira pessoa a desenvolver um método para reproduzir sons, e este conhecimento foi usado no desenvolvimento de outros importantes aparelhos tais como o toca-discos. Em 1882, Bell tornou-se um cidadão norte-americano. Ele faleceu em Baddeck, na Nova Escócia, em 1922.

1.2. NOÇÕES FUNDAMENTAIS DE ACÚSTICA

O som é a sensação causada no sistema nervoso pela vibração de membranas no ouvido, como resultado da energia gerada pela vibração de corpos quaisquer (tais como cordas de violão, cordas vocais e até mesmo o mar ao bater na areia da praia) dentro de uma faixa de freqüências específica transmitida através de um meio de propagação como, por exemplo, o ar ou a água.

Do ponto de vista físico, o som é a variação de pressão e conseqüente criação de uma onda que se propaga com uma certa velocidade através de um meio natural. Estas seriam portanto as principais variáveis relacionadas ao som no campo acústico.

Conforme citado acima, o ar ambiente constitui um meio pelo qual o som pode ser transmitido. Entretanto, outros meios, quer sólidos ou líquidos, podem servir para sua propagação. Constata-se que um meio com maior densidade, isto é, um sólido, propaga som melhor (com menos perdas) que o ar. Por exemplo, uma pessoa que encosta seu ouvido no trilho da linha férrea, pode constatar a presença de um trem a longa distância, mesmo sem conseguir ouvi-lo pelo ar.

1.2.1. Características da Onda Sonora e da Voz

Período e Freqüência – Cada oscilação de um corpo pode se repetir no tempo; o intervalo de tempo no qual uma onda se reproduz é chamado de período. A definição de freqüência é o número de vezes que essa onda se repete por segundo.

Amplitude – É a medida de quão afastada do referencial está a amostra do som num determinado instante. Comumente dá-se este nome ao maior destes valores.

Comprimento de Onda – é uma relação entre a velocidade de propagação do som no meio e sua freqüência. A fórmula a seguir explicita tal relação:

$$\lambda = \nu / f \tag{1}$$

As formas de onda podem ser caracterizadas como simples ou complexas. Uma forma de onda simples é constituída por uma única freqüência. Uma forma de onda complexa é composta por duas ou mais freqüências, neste caso a mais baixa será denominada de freqüência fundamental, tom principal ou primeiro harmônico, e representará a altura (o *pitch*)

do som; as demais serão chamadas de harmônicas e serão sempre múltiplos da fundamental (por exemplo, o 2° harmônico é o dobro da fundamental, o 3° é o triplo e assim por diante).

Também se classificam os sinais (as formas de onda) como determinísticos ou aleatórios. Enquanto que os determinísticos são geralmente descritos por uma equação conhecida (como uma senóide, por exemplo), os aleatórios têm seu comportamento caracterizado através de médias estatísticas.

Quando se analisa apenas uma onda senoidal pura com sua curva definida e sua freqüência fixa, diz-se que se trata de um "som puro". A voz, a música e os ruídos naturais são sons complexos, com características aleatórias e freqüências múltiplas.

Na fala (voz), dividem-se os sons em "sonoros" (ou "vibrantes") e surdos (ou "não-vibrantes"). No primeiro caso, os sons são restritos em faixa espectral de freqüência e resultam da vibração das cordas vocais propriamente ditas; já no segundo caso, têm espectro mais espalhado e são obtidos pela passagem forçada do ar através da região buconasal. A combinação deste dois tipos de sons forma a voz humana.

O mecanismo de produção da voz apresenta uma resposta limitada em freqüência. Esse limite é variável, mas fica por volta de 10 kHz. Os sistemas telefônicos limitam o sinal de voz a uma faixa de 3,4 kHz, com uma perda tolerável de qualidade.

1.2.2. Parâmetros do Som

Os parâmetros do som podem ser descritos através de algumas variáveis tais como a *velocidade de propagação* no meio. No caso do ar por exemplo temos a seguinte equação:

$$\nu = 331{,}4\sqrt{\frac{T}{273}} \qquad (2)$$

Onde T é a temperatura em Kelvin.

Outro parâmetro é o *espectro de freqüência*, que na faixa audível varia de 20 Hz a 20 kHz; para a grande maioria das pessoas essa faixa é reduzida, sendo que as freqüências abaixo de 20 Hz são denominadas de infra-som e as acima de 20 kHz são chamadas de ultra-som (muito utilizadas, por exemplo, na medicina para exames pré-natais); para aplicações musicais, onde a freqüência se chama altura, esse espectro varia de 20 Hz a 10kHz e para telefonia de 300 Hz a 3,4 kHz.

Ainda devemos considerar a *amplitude do som* que determina a sua intensidade e é função da potência com que um som é produzido por uma fonte. De acordo com a amplitude,

podemos classificar um som entre fraco ou forte. A amplitude de um som é determinada pela seguinte função: $I(t) = P^2(t) / \rho_0 . \nu$, onde P é a potência, ρ é a densidade do meio e ν é a velocidade do som no meio.

Por fim, podemos falar sobre o *Timbre*. Esta característica do som é fundamental para que distingamos os sons (e vozes) de mesma freqüência emitidos por fontes diferentes (tais como pessoas ou instrumentos musicais diferentes). O timbre de um determinado som é a combinação dos harmônicos da freqüência fundamental e suas amplitudes. Podemos assim entender porque uma nota musical qualquer, tocada em dois instrumentos distintos, pode ser diferenciada.

1.3. INTRODUÇÃO AO SISTEMA TELEFÔNICO

1.3.1. Aparelho Telefônico

O aparelho telefônico é o responsável pela origem e recepção das ligações. Apesar de seu aspecto simples, ele desempenha um grande número de operações. Suas funções incluem:

- Solicitação para o uso do sistema telefônico, quando o monofone é levantado
- Indicar que o sistema está pronto para uso, por meio da recepção do tom de discar
- Enviar o número do telefone chamado ao sistema
- Indicar o estado da ligação, por meio de sinalização acústica
- Acusar o recebimento de uma ligação, com o toque da campainha
- Converter a voz em sinais elétricos para a transmissão
- Ajustar automaticamente a mudança de potência
- Sinalizar ao sistema o término de uma ligação.

A figura abaixo ilustra o um telefone com seus principais componentes. Os telefones funcionam com tensão contínua de – 48 V (quando "no gancho"), corrente de operação de 20 a 80 mA , perda típica de enlace de 8 dB e distorção de – 50 dB.

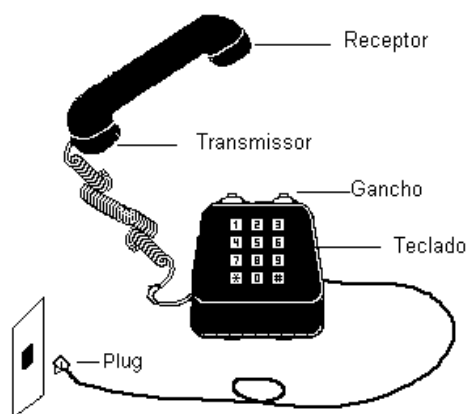



Figura 1-3.1: Principais componentes de um telefone

O telefone decádico, no qual os dígitos são transmitidos por seqüências de pulsos, está com seus dias contados. A figura abaixo mostra uma versão do teclado de um telefone multifreqüencial, no qual os dígitos são transmitidos por combinações de freqüências, com

um par de freqüências associado a cada tecla. O sistema de discagem multifreqüencial está substituindo o telefone decádico por apresentar as seguintes vantagens:

- Diminui o tempo de discagem
- Utiliza componentes eletrônicos de estado sólido
- Pode ser usado para a transmissão de dados a baixas taxas
- Reduz os requisitos de equipamentos na central local
- É mais compatível com as Centrais de Programa Armazenado (CPA)

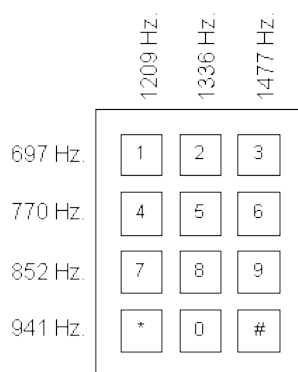



Figura 1-3.2: Teclado de um telefone multifreqüencial

O fundamento da telefonia é o estabelecimento da ligação telefônica. Para tanto, além do telefone e do sistema telefônico, é necessária a existência do assinante. O processo inicia-se com o desejo de um determinado assinante A de conversar com o outro assinante, digamos B. O assinante A começa então uma chamada por meio do sistema telefônico. Essa chamada pode ser atendida pelo assinante B, ou sofrer alguma interrupção por conta de congestionamento no sistema, erro na discagem, telefone ocupado ou ausência de resposta por parte do assinante chamado.

Em função do sinal recebido, o assinante A pode tomar a decisão de desistir, ou renovar a tentativa. Essa atitude pode ser tomada imediatamente, ou depois de algum tempo. A nova tentativa irá, dessa forma, ocupar novamente o sistema telefônico.

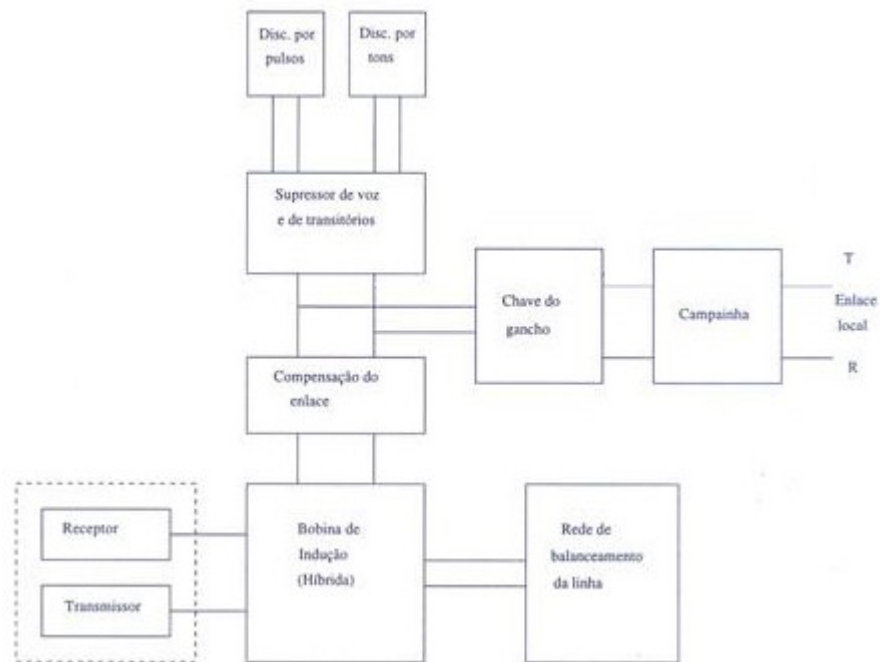



Figura 1-3.3: Diagrama de blocos do telefone

1.3.2. Distorções introduzidas pelo Sistema Telefônico

O projeto de um sistema de telefonia digital deve levar em conta todos os aspectos da rede, do locutor ao ouvinte. Algumas características dos sistemas telefônicos levam à distorção no sinal de voz. A lista que segue ilustra alguns dos problemas encontrados e seus efeitos sobre a inteligibilidade:

- Limitação na amplitude de pico do sinal - Afeta a qualidade da voz, mas não reduz apreciavelmente a inteligibilidade quando a fala é ouvida em ambiente silencioso e sob índices de percepção confortáveis.
- Corte central no sinal - A supressão dos níveis mais baixos do sinal causa um efeito drástico sobre a inteligibilidade do sinal e afeta a qualidade da fala.
- Deslocamento de freqüência - Ocorre quando a freqüência recebida difere da transmitida e afeta a inteligibilidade e o reconhecimento do locutor.
- Retardo em sistemas operados por voz - Resulta na omissão da parcela inicial de uma mensagem. Afeta a inteligibilidade com uma queda linear da mesma com o aumento do intervalo omitido.
- Defasagem e retardo de transmissão - é normalmente mais pronunciado na transmissão via satélite ou de longa distância, por conta da distância que o sinal

portador tem que percorrer. Como a inteligibilidade é resistente ao retardo, este último afeta principalmente a qualidade da fala. Circuitos supressores de eco acabam eliminando parte da sílaba inicial, em transmissões via satélite.

- Eco - Resulta de reflexões do sinal em pontos terminais da linha. Retardos acima de 65 ms produzem ecos perceptíveis e retardos inferiores tendem a tornar o som deturpado.
- Realimentação - Realimentação acústica pode ocorrer em trajetos de redes complexas. O efeito é perturbador para o locutor e para o ouvinte.
- Ruído - Diversos tipos de ruído afetam a transmissão do sinal de voz. O ruído básico para sistemas digitais, conhecido como ruído do quantização, resulta do mapeamento do sinal analógico em digital. O ruído é um sinal aleatório por natureza e provoca uma sensação desagradável ao ouvido, devendo ser minimizado na medida do possível.

Redes Telefônicas

2.1. INTRODUÇÃO

A rede telefônica evoluiu a partir do serviço telefônico básico para uma oferta variada de serviços, tornando-se uma estrutura de comunicações complexa e de alta capilaridade, e atingindo mais de um bilhão de linhas pelo mundo no ano de 2001.

A rede telefônica é composta por:

- Redes de longa distância, que incluem centrais internacionais e interurbanas e seus respectivos entroncamentos;
- Redes locais, contendo as centrais e entroncamentos em área urbana;
- Enlace do assinante, constituído pelos terminais e linhas de assinantes.

Quanto ao método de comutação, as redes podem utilizar a comutação por circuitos ou por pacotes. Na comutação por circuitos os recursos necessários em todos os subsistemas de telecomunicações que ligam origem e destino, são ocupados enquanto durar a conexão. Esta estratégia é adequada para chamadas com alto grau de utilização do meio. A comutação por circuitos ainda é a mais utilizada.

A comutação por pacotes é apropriada para sistemas com fator de utilização mais baixo, nos quais os recursos são utilizados apenas por uma fração do tempo, e é empregada principalmente na comunicação entre computadores. Pode-se prever a sua utilização em telefonia num futuro próximo.

Um site fundamental para a consulta de normas aplicadas pelo Sistema Telebrás (assumidas pela Anatel) é:

Sistemas.anatel.gov.br/sdt

Consultar práticas por série, série Engenharia

2.2. A CENTRAL TELEFÔNICA

A central telefônica é o elemento de rede responsável pela comutação de sinais entre os usuários, automatizando o trabalho das antigas telefonistas que comutavam manualmente os caminhos para a formação dos circuitos telefônicos. As linhas telefônicas dos vários assinantes chegam às centrais telefônicas e são conectadas entre si, estabelecendo circuitos temporários que permitem o compartilhamento de meios, promovendo uma otimização dos recursos disponíveis.

A comutação era eletromecânica até o inicio dos anos 70, quando as funções lógicas de comando e controle da comutação passaram a ser executadas por dispositivos eletrônicos. A conexão continuou eletromecânica. Somente na década de 80 a comutação passou a ser totalmente eletrônica. Essas centrais empregam computadores para a gestão de processos e são conhecidas como Centrais de Programa Armazenado (CPA's).

A central a que estão conectados os assinantes de uma rede telefônica em uma região é chamada de ***Central Local***. Para permitir que assinantes ligados a uma Central Local falem com os assinantes ligados a outra Central Local são estabelecidas conexões entre as duas centrais, conhecidas como circuitos troncos. No Brasil um circuito tronco utiliza geralmente o padrão internacional da UIT para canalização digital sendo igual a 2 Mbps ou E1.

Em uma cidade podemos ter uma ou várias Centrais Locais. Em uma região metropolitana pode ser necessário o uso de uma ***Central Tandem*** que está conectada apenas a outras centrais, para otimizar o encaminhamento do tráfego. As centrais denominadas ***Mixtas*** possuem a função local e a função tandem simultaneamente.

Estas centrais telefônicas locais estão também interligadas a Centrais Locais de outras cidades, estados ou países através de centrais de comutação intermediárias denominadas de ***Centrais Trânsito***. As Centrais Trânsito são organizadas em classes conforme sua área de abrangência, sendo as Centrais Trânsito Internacionais as de mais alta hierarquia. É possível desta forma conectar um assinante com outro em qualquer parte do mundo. A topologia de uma rede telefônica é ilustrada na figura abaixo.

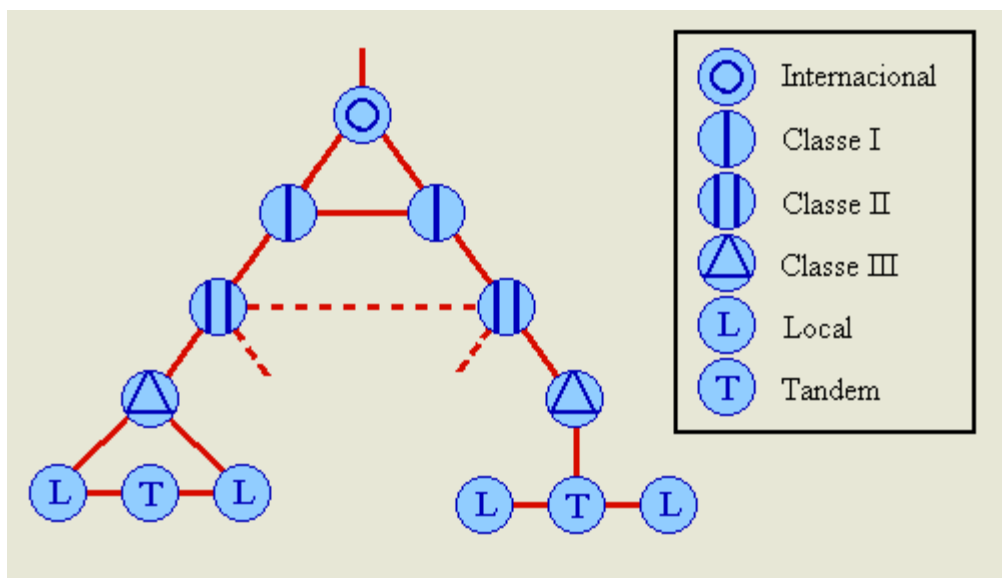



Figura 2-2.1: Topologia de uma rede telefônica

2.3. SINALIZAÇÃO NA REDE TELEFÔNICA

Para que uma chamada seja estabelecida o sistema telefônico tem que receber do assinante o número completo a ser chamado, estabelecer o caminho para a chamada e avisar ao assinante destinatário que existe uma chamada para ele. O sistema que cumpre estas funções em uma rede telefônica é chamado de sinalização. Um exemplo de sinalização no estabelecimento de uma ligação é ilustrado na figura abaixo.

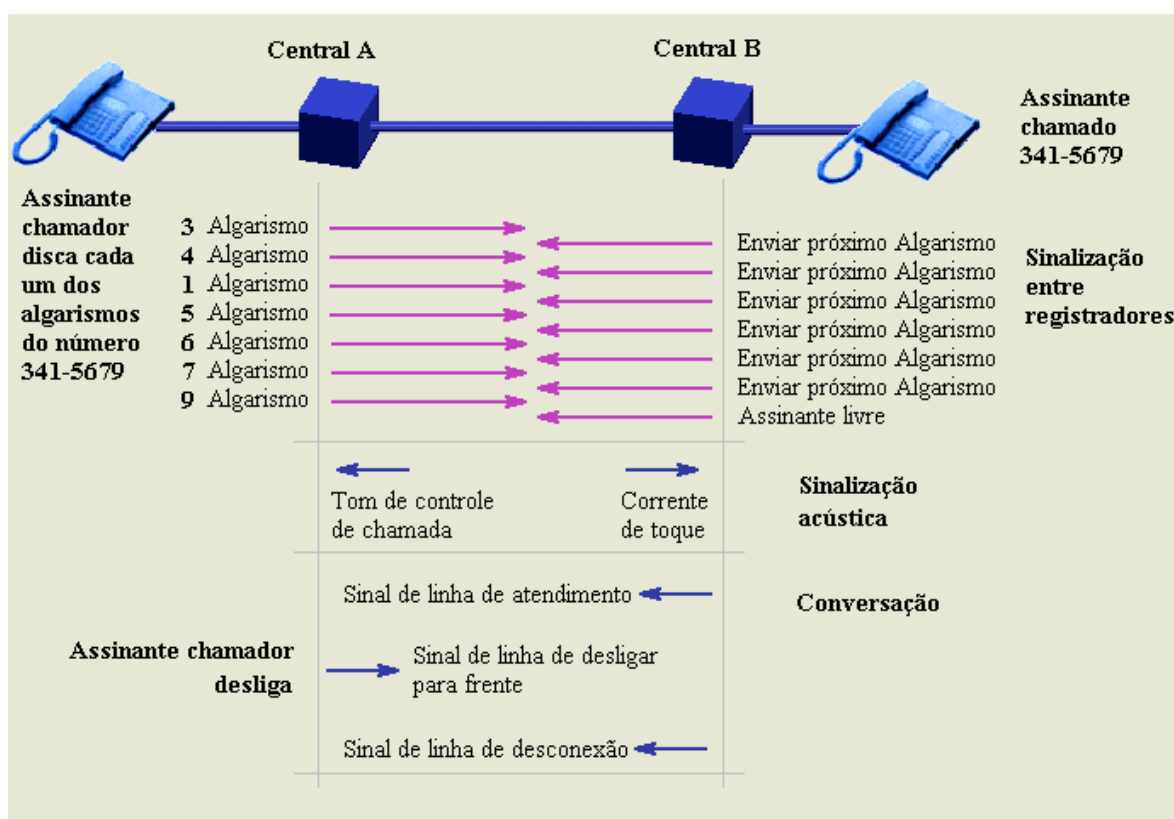



Figura 2-3.1: Exemplo de sinalização

2.3.1. Sinalização Acústica

A sinalização acústica é transmitida pelos órgãos da central de comutação diretamente aos assinantes sob a forma sonora, e tem por finalidade indicar o estado de operação dos sistemas telefônicos. Esta é a única sinalização perceptível pelos assinantes. Os sinais são os seguintes:

- O *Tom de discar* é a sinalização enviada pela central ao assinante chamador, indicando que a mesma está pronta para receber e armazenar os números teclados.

- O *Tom de controle de chamada* é enviado pela central indicando ao chamador que o usuário de destino está sendo chamado. Este sinal é enviado juntamente com a corrente de toque que vai para o assinante chamado.
- O *Tom de ocupado* é enviado ao assinante chamador, indicando uma das seguintes ocorrências: o assinante chamado está ocupado; há congestionamento em algum ponto da rede; o chamador não observou as regras de discagem; o número discado não está acessível à categoria do chamador; o terminal que retém a chamada desligou. Este sinal é gerado pela central do assinante chamador.
- O *Tom de número inacessível* é enviado ao assinante chamador para indicar que a chamada não pode ser completada por uma das seguintes razões: o número chamado não existe; a linha do assinante chamado está com defeito; o número do assinante mudou. Em algumas situações, este tom é substituído por uma mensagem gravada.
- O *Tom de aviso de chamada em espera* é o sinal enviado por uma central aos terminais envolvidos em uma conversação, ou apenas ao terminal chamado que dispõe do serviço "chamada em espera", indicando a existência de outra chamada. O assinante chamador em espera receberá o tom de controle de chamada enquanto este sinal é enviado.
- O *Tom de aviso de programação* é o sinal enviado ao terminal chamador, em substituição ao tom de discar, informando que o recebimento de tráfego está inibido por programação.
- A *Corrente de toque é a sinalização* enviada pela central ao assinante chamado, para indicar que há chamada para o mesmo. O sinal aciona a campainha do aparelho telefônico a intervalos iguais aos do tom de controle de chamada, porém eles podem não estar sincronizados.

Tom de discar	425 ± 25 Hz emitido continuamente.
Tom de controle de chamada	425 ± 25 Hz emitido durante 1 ± 0,1 s, seguido de um período de silêncio de 4 ± 0,4 s.
Tom de ocupado	425 ± 25 Hz emitido em intervalos de 250 ± 25 ms intercalados com intervalos iguais de silêncio.
Tom de número inacessível	425 ± 25 Hz emitido em períodos alternados de 250 ± 25 ms e 750 ± 75 ms intercalados com intervalos de silêncio de 250 ± 25 ms.
Tom de aviso de chamada em espera	425 ± 25 Hz emitido durante 50 ± 10 ms, seguido de um período de silêncio de 500 ± 100 ms.
Tom de aviso de programação	425 ± 25 Hz emitido em intervalos de 125 ms intercalados com intervalos iguais de silêncio.
Corrente de toque	25 ± 2,5 Hz emitido durante 1 ± 0,1 s, seguido de um período de silêncio de 4 ± 0,4 s.

Tabela 2-3.1: Características da sinalização acústica

2.3.2. Sinalização de Linha

Esta sinalização é responsável por efetuar a supervisão dos enlaces dos circuitos que interligam duas centrais, trocando informações relacionadas aos estágios da conexão, e agindo durante toda a conexão sem ser percebida pelos assinantes. Também é responsável por enviar os pulsos de tarifação, quando necessário.

Os sinais que são gerados no lado do assinante que origina a chamada são denominados Sinais para Frente, enquanto os gerados no lado do assinante chamado são os Sinais para Trás.

Os Sinais para frente são:

- Ocupação – é emitido pela central de onde provém a chamada para levar o circuito associado à condição de ocupação.
- Desligar para frente – é emitido pela central do assinante chamador no instante em que este repõe o telefone no gancho, para indicar que o chamador desligou, liberando a central de destino e todos os órgãos envolvidos na chamada.
- Re-chamada – O sinal de re-chamada ocorre geralmente quando se utiliza mesa operadora, para re-chamar o assinante chamado, após este ter desligado.

E os Sinais para trás:

- Atendimento – é gerado pela central para onde foi enviado o sinal de ocupação, indicando ao chamador o momento em que o assinante chamado atende a ligação.
- Desligar para trás – é enviado ao chamador indicando que o assinante chamado desligou.
- Confirmação de desconexão – é enviado em resposta a um sinal de desligar para frente, indicando que ocorreu a liberação dos órgãos associados à ligação.
- Desconexão forçada – é gerado após uma temporização pela central responsável pela tarifação, quando o usuário chamado desliga mas o chamador não. Sua temporização tem início no momento do envio da sinalização de desligar para trás, e geralmente é de 90 segundos.
- Tarifação – O sinal de tarifação é emitido a partir do ponto de tarifação para o contador do assinante chamador, de acordo com o degrau tarifário correspondente.
- Bloqueio – O sinal de bloqueio ocorre quando há falha ou bloqueio (efetuado por operador) na central do assinante chamado.

2.3.3. Sinalização de Registrador

A sinalização de registrador corresponde ao conjunto de sinais responsáveis pela troca de informações destinadas ao estabelecimento das chamadas (número do assinante chamador, categoria do assinante chamador, etc.).

Esta sinalização é trocada entre órgãos de controle das centrais, ocorrendo no início da ligação, entre assinantes de centrais distintas, até o momento em que o assinante chamador ouve o sinal sonoro indicando que o outro assinante está sendo chamado, está ocupado ou não existe.

A sinalização entre registradores pode ser por pulsos decádicos ou por sinais multifreqüenciais, sendo que esta última divide-se em MF (multifreqüencial) ou MFC (multifreqüencial compelida). As centrais CPA modernas adotam apenas a sinalização MFC.

Sinalização MFC

Utiliza-se o nome sinalização compelida, pois na sinalização MFC cada sinal enviado compele o registrador de destino a emitir um sinal de volta, caso contrário a ligação é interrompida. Assim, a duração de um sinal é determinada pela recepção de outro sinal enviado no sentido oposto como resposta ao primeiro. Por exemplo, um sinal para frente

permanece sendo emitido enquanto não é recebido, em contrapartida, um sinal para trás, ou não é limitado por um temporizador.

Nº	Freq. (Hz)	Grupo I	Grupo II
1	1020 e 1140	Algarismo 1	Assinante comum
2	900 e 1140	Algarismo 2	Assinante tarifação imediata
3	900 e 1020	Algarismo 3	Equipamento de teste
4	780 e 1140	Algarismo 4	Telefone público
5	780 e 1020	Algarismo 5	Mesa operadora
6	780 e 900	Algarismo 6	Equipamento de transmissão de dados
7	660 e 1140	Algarismo 7	Telefone público interurbano
8	660 e 1020	Algarismo 8	Comunicação de dados - serviço internacional
9	660 e 900	Algarismo 9	Assinante com prioridade - serviço internacional
10	660 e 780	Algarismo 0	Telefonista com facilidades de transferência - serviço internacional
11	540 e 1140	Inserção de semi-supressor de eco na origem	Assinante com facilidade de transferência
12	540 e 1020	Pedido recusado ou indicação de trânsito internacional	Reserva
13	540 e 900	Acesso a equipamento de teste	Reserva
14	540 e 780	Inserção de semi-supressor de eco ou indicação de trânsito internacional	Reserva
15	540 e 660	Fim de número ou indicação de que a chamada cursou enlace via satélite	Reserva

Tabela 2-3.2: Sinais para frente

Os sinais MFC são formados por combinações de duas freqüências dentre dois blocos de seis, resultando em 15 combinações possíveis para cada bloco (sinais para frente e para trás). Como a faixa de freqüências vocais estende-se de 300 a 3400 Hz, e os extremos desta faixa são um pouco mais atenuados que o restante, escolheu-se a referência de 1260 Hz a partir da qual são alocadas as freqüências, com espaçamento de 120 Hz As freqüências maiores que 1260 Hz são utilizadas pelos sinais para frente, e as menores pelos sinais para trás.

Nº	Freq. (Hz)	Grupo A	Grupo B
1	1020 e 1140	Enviar o próximo algarismo	Assinante livre com tarifação
2	900 e 1140	Enviar o primeiro algarismo	Assinante ocupado
3	900 e 1020	Passar para o grupo B	Assinante com número mudado
4	780 e 1140	Congestionamento	Congestionamento
5	780 e 1020	Enviar categoria e identidade do chamador	Assinante livre sem tarifação do chamador

6	780 e 900	Reserva	Assinante livre com tarifação e retenção sob o controle do assinante chamado.
7	660 e 1140	Enviar o algarismo N-2	Número vago
8	660 e 1020	Enviar o algarismo N-3	Assinante com defeito
9	660 e 900	Enviar o algarismo N-1	Reserva
10	660 e 780	Reserva	Reserva
11	540 e 1140	Enviar a indicação de trânsito internacional	Reserva internacional
12	540 e 1020	Enviar digito de idioma ou discriminação	Reserva
13	540 e 900	Enviar indicador do local do registrador internacional de origem	Reserva
14	540 e 780	Solicitar informações da necessidade de semi-supressor de eco no destino	Reserva
15	540 e 660	Congestionamento na central internacional	Reserva

Tabela 2-3.3: Sinais para trás

Os sinais para frente utilizam uma combinação de duas das freqüências 1380 Hz, 1500 Hz, 1620Hz, 1740 Hz, 1860 Hz e 1980 Hz. Estes sinais são divididos em Grupo I, referente a informações numéricas e de seleção, e Grupo II, referente a informações sobre categoria do assinante chamador. O sinal recebido é interpretado com sendo do Grupo I ou II de acordo com o contexto da sinalização. Por exemplo, se a central destino enviou um sinal para trás solicitando a categoria do chamador, o próximo sinal para frente recebido será interpretado como Grupo II (categoria do assinante). A tabela 2-3.2 apresenta os sinais para frente em detalhes.

Os sinais para trás utilizam as freqüências 540 Hz, 660 Hz, 780 Hz, 900 Hz, 1020 Hz e 1140 Hz. Estes sinais são classificados em Grupo A, que se refere a solicitações para possibilitar o estabelecimento da conexão, e Grupo B, referente ao estado e à categoria do assinante chamado. A passagem de um grupo para outro é determinada pela central destino. A tabela 2.3.3 apresenta os sinais para trás em detalhes.

2.4. SINALIZAÇÃO POR CANAL COMUM

Neste tipo de sistema, a sinalização entre centrais utiliza canais exclusivamente dedicados à sinalização, isto é, independentes dos canais de voz, que são responsáveis pela troca de informações relativas a todas as chamadas em andamento ou em estabelecimento. Como o tempo gasto para a sinalização é relativamente curto quando comparado com o tempo de conversação, a sinalização por canal comum apresenta a grande vantagem de conseguir em um único canal de sinalização, tratar de milhares de chamadas, além da flexibilidade de inserir novos sinais com o aparecimento de novos serviços.

Pode-se dizer que entre duas centrais que se comunicam utilizando o sistema de sinalização por canal comum, existe um canal de dados responsável pela transmissão da sinalização correspondente aos demais canais (na prática pode haver mais de um).

O ITU-T padronizou um sistema de sinalização por canal comum chamado Sistema de Sinalização Número 7, ou SS#7, que é o sistema adotado no Brasil. Como o canal de sinalização é um canal de dados entre as centrais, o SS#7 padroniza um protocolo de comunicação digital, baseado no modelo OSI (*Open System Interconnection*). É importante observar que o canal de sinalização, por definição, não precisa utilizar o mesmo caminho dos canais de áudio. Entretanto, para aproveitar a rede instalada, o SS#7 padronizou a utilização de canais de sinalização com taxa de 64kbit/s de forma que este canal possa ser um dos canais de um tronco digital de 2Mbit/s, transportando informações de sinalização referentes aos canais de áudio, mesmo daqueles que não estiverem presentes no mesmo tronco digital.

O canal de sinalização pode ocupar qualquer um dos canais do tronco digital exceto o canal 0 (zero) que transporta informação de sincronismo. Normalmente é utilizado o canal 16, mas ele não precisa, necessariamente, ser utilizado como canal de sinalização, ou seja, caso exista mais de um tronco de 2Mbit/s operando entre duas centrais, a sinalização poderá utilizar um canal de um dos troncos e transmitir informações de todos os canais de áudio de todos os troncos.

Quando as mensagens entre duas centrais são transportadas em uma rota de sinalização que consiste de um enlace direto entre duas centrais, a sinalização é dita de Modo Associado. Se a rota de sinalização entre as duas centrais é composta por mais de um enlace de sinalização (SL - *Signaling Link*), ela é dita Modo Não-Associado. Neste caso, um ou mais pontos de transferência de sinalização (STP) são usados para a transferência do tráfego de sinalização.

2.4.1. Arquitetura da Rede

Na rede de SS#7, os sinais são transferidos utilizando a comutação de pacotes, isto é, as informações de sinalização são agrupadas em pacotes que são transferidos utilizando um canal de 64kbits/s. Por ser uma rede independente da rede de telefonia, é necessário definirmos os conceitos principais da rede de sinalização SS#7.

Cada componente da rede de SS#7 é chamado de ponto de sinalização. Assim, cada central da rede é um ponto de sinalização, e existem 3 funções distintas que podem ser executadas pelos pontos de sinalização, que os definem. O primeiro ponto é aquele responsável pelo envio e pela recepção das informações, ou seja, é o ponto correspondente às centrais de comutação telefônica. O segundo ponto é responsável pelo roteamento ou transferência das informações de sinalização em direção ao seu destino. Portanto eles não originam mensagens, servindo apenas de roteadores para as mensagens entre origem e destino. O último ponto ou função, é permitir o acesso a bancos de dados. Como a SS#7 é digital, é razoável que a mesma possa permitir que os pontos de sinalização tenham acesso a bancos de dados centralizados, com o objetivo de prover serviços mais elaborados que simplesmente os serviços telefônicos.

Os três tipos de pontos de sinalização referidos anteriormente, são definidos como:

- *Service Switchig Point* (SSP) ou Ponto de Serviço (PS): corresponde às centrais de comutação. São estas centrais que geram as mensagens de sinalização telefônica que devem ser transmitidas de um SSP para um outro SSP.
- *Signal Transfer Point* (STP) ou Ponto de Transferência de Sinalização (PTS): responsável pelo roteamento das mensagens de sinalização entre os SSP's. Não tem função de comutação de áudio, embora muitos equipamentos possam executar tanto a função de STP como de SSP.
- *Service Control Point* (SCP): corresponde aos bancos de dados que podem ser acessados pelos demais pontos da rede para obter informações necessárias para a disponibilização de serviços mais elaborados.

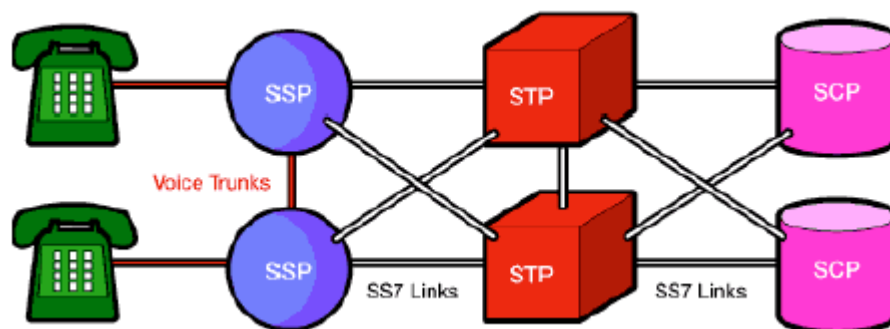



Figura 2-4.1: Arquitetura da Rede

Cada ponto da rede de sinalização possui um endereço chamado *point code*, e é através deste *point code* (endereço) que um ponto da rede consegue acessar um outro ponto. Isto é realizado inserindo, em cada mensagem enviada, o endereço correspondente ao ponto de destino que se deseja acessar.

2.4.2. Blocos funcionais do SS#7

Na aplicação do modelo OSI (*Open System Interconection*, apresentado no capítulo referente a redes) ao caso do sistema telefônico, os assinantes e o aparelho telefônico são responsáveis pelas funções dos níveis 4 a 7. A rede telefônica se encarrega dos níveis 1 a 3. O sistema telefônico é composto dos seguintes blocos funcionais:

- MTP (*Message transfer part*) Nível 1 – Funções de enlace de dados. Definem as características físicas, elétricas e funcionais do enlace de dados;
- MTP Nível 2 – Funções de enlace de sinalização. Definem os procedimentos para transferência confiável de mensagens de sinalização. A mensagem, com comprimento variável, tem campos de controle para confiabilidade, além de informação e sinalização;
- MTP Nível 3 – Funções de rede de sinalização. Estabelecem os procedimentos para controlar o encaminhamento das mensagens. No caso de falhas, controla também as reconfigurações para preservar ou restabelecer a capacidade normal de transferência de mensagens;
- SCCP (*Signalling connection control part*) – Provê funções adicionais à MTP, que completam a camada 3[1] do modelo OSI, para fornecer serviços orientados ou não a

[1] Sobre as camadas do modelo OSI, ver item 3.3

conexões. Incluem a tradução de dígitos discados em códigos do ponto de sinalização, que permitem o roteamento das mensagens.

- TUP (*Telephone user part*) – Define as funções e procedimentos de sinalização necessários ao uso do SS#7 no controle da chamada telefônica;
- ISUP (*ISDN user part*) - Define as funções e procedimentos necessários ao oferecimento de serviços comutados e facilidades de usuários para aplicações de voz e dados na RDSI;
- TC (*Transaction capability*) – Indica os serviços e protocolos da camada de aplicação. Divide-se em TCAP (*transaction capability application part*) e ISP (*intermediate service part*);
- AE (*Application entity*) – Representa as funções de comunicação de um processo de aplicação, sendo composta de um conjunto de ASE (*application service elements*).

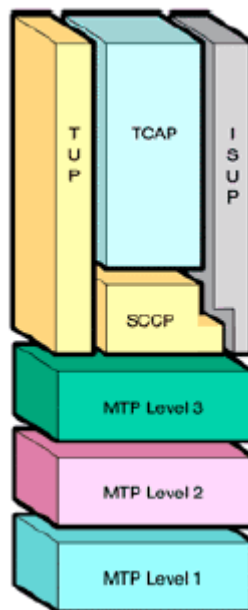



Figura 2-4.2: Blocos Funcionais

2.5. OPERAÇÃO PLESIÓCRONA

Uma das dificuldades práticas da transmissão por TDM é o fato de que os multiplexadores trabalham com sinais gerados por equipamentos cuja taxa de transmissão pode ser ligeiramente diferente (chamados de sinais plesiócronos do grego *plesio*, quase igual), e, para a ocorrência da multiplexação, é essencial que todos os bits de entrada estejam sincronizados.

Desta forma, antes da etapa de multiplexação faz-se necessário o ajuste das taxas de transmissão, feito pela adição de bits sem informação, denominados bits de justificação. No processo de demultiplexação estes bits são identificados e descartados, regenerando o sinal original.

Este problema de sincronização ocorre em todos os níveis da hierarquia TDM, de forma que em cada estágio há a adição de bits de justificação. Por este motivo adotou-se o nome de Hierarquia Digital Plesiócrona (PDH, *Plesiochronous Digital Hierarchy*) para esta linha de multiplexadores TDM.

A tabela 2-5.1 mostra a hierarquia PDH utilizada na América do Norte e Europa. O padrão PDH Europeu, recomendado pelo ITU (norma G.702), é adotado na maior parte do mundo, inclusive no Brasil.

Hierarquia	América do Norte		Europa	
0	DS0	64 kbit/s	E0	64 kbit/s
1	DS1 ou T1	1544 kbit/s	E1	2048 kbit/s
2	DS2 ou T2	6132 kbit/s	E2	8448 kbit/s
3	DS3 ou T3	32064 kbit/s	E3	34368 kbit/s
4	DS4	97728 kbit/s	E4	139264 kbit/s

Tabela 2-5.1: Hierarquia PDH na América do Norte e Europa

A redução dos custos dos circuitos integrados na transmissão por fibra ótica tornou as redes PDH baratas, o que levou à multiplicação deste sistema por todo o mundo. Contudo, o PDH foi concebido visando atender simplesmente transmissão de voz, e a sua flexibilidade para atender a demanda de novos serviços com taxas e qualidade crescentes não era suficiente, devido ao processamento de bits necessário.

A evolução dos padrões para sistemas de transmissão levou ao desenvolvimento do *Synchronous Digital Hierarchy*, um sistema totalmente síncrono que vem substituindo gradativamente o PDH.

2.6. MONTAGEM DE REDE TELEFÔNICA DIGITALIZADA

A comunicação entre as centrais CPA-T é feita através de um sinal digital TDM-PCM de primeira ordem. Para garantir a integridade na transmissão, algumas adequações são feitas.

Nas redes locais (centrais e entroncamentos em área urbana) é usada a codificação de fonte, equalização e filtragem do sinal para corrigir distorção e interferência intersimbólica. Para grandes distâncias, há o uso de modulação digital. Alguns tipos de sinais utilizados na transmissão em banda básica entre centrais são mostrados na figura abaixo.

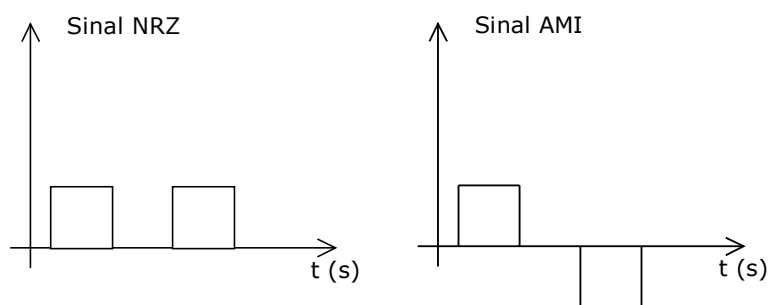



Figura 2-6.1: Tipos de sinais utilizados na transmissão em banda básica entre centrais.

Nas redes de longa distância (centrais internacionais e interurbanas e os respectivos entroncamentos) os sinais são multiplexados e transmitidos em banda larga. Os meios de transmissão usados são: transmissão por cabo, rádio digital, fibra ótica e transmissão via satélite.

Para os serviços digitais absorvidos pela rede telefônica, é necessário utilizar meios de alta capacidade, TDM's de ordem elevada e hierarquia digital síncrona (SDH).

2.6.1. Codificação de Linha

O pulso básico de transmissão digital (Figura 2-6.2) é o pulso retangular ou função porta, representado pela equação: $p_T(t) = \mathrm{A}\,[\mathrm{u}(t + T/2) - \mathrm{u}(t - T/2)]$ ou

$$p_T(t) = \begin{cases} A \text{ se } |t| \le \dfrac{T}{2} \\ 0 \text{ se } |t| > \dfrac{T}{2} \end{cases},$$

em que u(t) é a função degrau unitário dada por:

$$u(t) = \begin{cases} 1 \text{ se } t \geq 0 \\ 0 \text{ se } t < 0 \end{cases}.$$

A transformada de Fourier do pulso básico é dada por:

$$F(\omega) = \int_{-T/2}^{T/2} Ae^{-j\omega t} dt = \frac{A}{j\omega}\left(e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}}\right) = \frac{A}{j\omega} 2j \operatorname{sen} \frac{\omega T}{2},$$

que pode ser escrita como:

$$F(\omega) = AT\left(\frac{\operatorname{sen}(\omega T/2)}{(\omega T/2)}\right) \text{ e, finalmente, } F(\omega) = AT\,Sa\left(\frac{\omega T}{2}\right),$$

em que $Sa(x) = \frac{sen(x)}{x}$. Essa última função é conhecida como função amostragem ou sampling, cujo módulo está ilustrado na Figura 2-6.3.

A banda passante para a transmissão do pulso pode ser obtida considerando-se o primeiro cruzamento em zero do espectro, ou seja,

$$B = \frac{1}{T} \text{ Hz}.$$

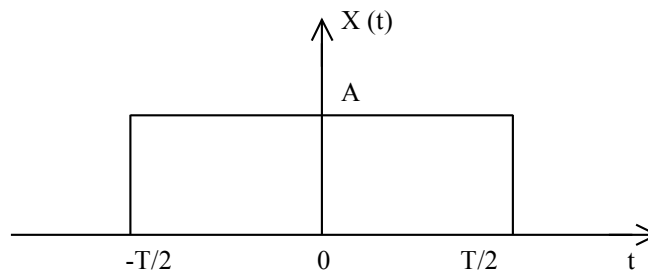



Figura 2-6.2: Pulso básico de transmissão

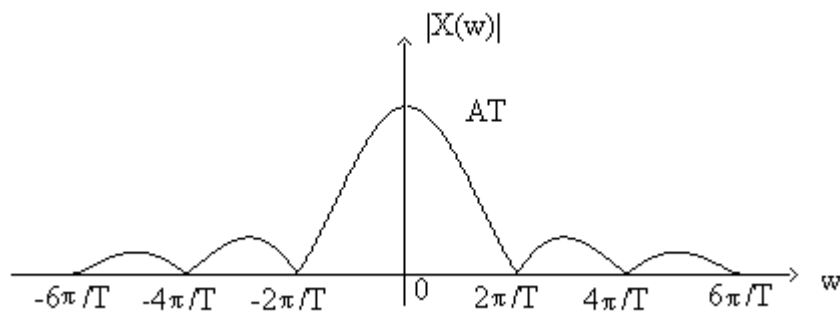



Figura 2-6.3: Espectro do pulso básico de transmissão

Conforme ilustrado na figura 2-6.3, o espectro do sinal espalha-se em uma faixa bem maior do que a banda passante nominal. Isto pode ocasionar interferência em pulsos transmitidos por portadoras distintas. A codificação de linha, no entanto, oferece maneiras de adequar o pulso à transmissão.

Os sinais transmitidos estão sujeitos ao ruído e à interferência intersimbólica. As principais causas dessas perturbações são: imprecisão temporal, banda passante insuficiente, distorção de amplitude e distorção de fase.

A solução para permitir a recuperação dos símbolos transmitidos é a utilização de técnicas de codificação de linha como: utilização de códigos específicos, uso de bits de temporização, embaralhamento dos dados, etc. Os códigos de linha mais utilizados são (Alencar, 1998):

- NRZ (Non-Return to Zero), ou código polar;
- AMI (Alternate Mark Inversion), código bipolar em que os pulsos têm a polaridade invertida;
- BNZS (Binary N Zero Substitution), que substitui seqüências determinadas de zeros por uma seqüência especial de pulsos para produzir violação no código AMI, exemplo B3ZS;
- Codificação ternária, que implica na codificação de grupos de 4 bits em três dígitos ternários (4B3T);
- Código Manchester ou bifase, usado no padrão Ethernet IEEE 802.3 para redes locais;
- CMI (Coded Mark Inversion), padronizado pelo CCITT, no qual a técnica AMI é conhecida com a codificação dos zeros por ondas quadradas de meio-ciclo e determinada fase.

2.6.2. Transmissão entre Centrais CPA-T

A transmissão de informação é feita através de um meio de propagação, que pode ser não guiado como a atmosfera, para a transmissão de rádio, por exemplo, ou guiado como as linhas telefônicas.

2.6.3. Transmissão via rádio

No caso de radiotransmissores, para que um sinal possa ser irradiado com eficácia, a antena irradiadora deve ser da ordem de um décimo ou mais do comprimento de onda correspondente à freqüência do sinal a ser transmitido. Desta forma, para transmitir sinais de voz, as antenas necessitariam de dimensões da ordem de centenas de quilômetros.

Para tornar possível a transmissão de sinais com a freqüência da voz humana, este deve ser modulado. A modulação utiliza ondas portadoras que servem como um suporte para levar a informação, também chamado sinal modulador. A onda senoidal é usada tradicionalmente como portadora, com a modulação podendo se processar geralmente de três maneiras (Alencar, 1998):

- Modulação em Amplitude (AM), quando o parâmetro da portadora a ser variado é a amplitude. Inclui os sistemas ASK, OOK e ASK M-ário;
- Modulação em Ângulo, quando o parâmetro da portadora a ser variado é a sua fase (PM) ou freqüência (FM). Os sistemas mais comuns são o FSK, PSK e PSK M-ário;
- Modulação em Quadratura (QUAM) quando são variados os parâmetros de amplitude e ângulo ao mesmo tempo. Dentre os mais conhecidos estão o QAM e o QPSK.

A figura 2-6.4 mostra um exemplo de transmissão via rádio digital.

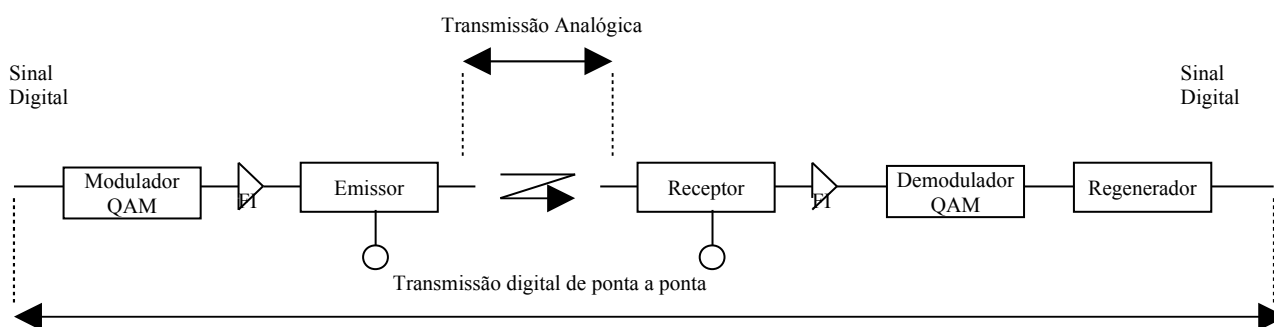



Figura 2-6.4: Transmissão via rádio digital

~

2.6.4. Transmissão por Canal Guiado

O canal guiado permite o confinamento da onda eletromagnética do transmissor ao receptor através do guia, sendo por isso adequado para áreas de campo eletromagnético intenso. Além disso, possibilita contornar obstáculos e facilita as aplicações em enlaces privados de comunicações.

Porém, há necessidade de casamento de impedâncias para evitar perdas por reflexão e possui um custo de implementação mais elevado que os sistemas de rádio, principalmente no caso de longas distâncias e em centros urbanos.

O padrão adotado atualmente para transmissão guiada entre centrais telefônicas é o canal de fibra ótica, em virtude do grande volume de informação trafegada. O cabo coaxial é outro tipo de canal guiado com aplicações em telefonia, mas sua aplicação mais comum é em redes locais de computadores.

A transmissão guiada entre centrais por meio de cabo óptico é ilustrada na figura abaixo. A partir da década de 80, esse tem sido o meio preferencial para fazer o entroncamento entre as centrais do sistema telefônico.

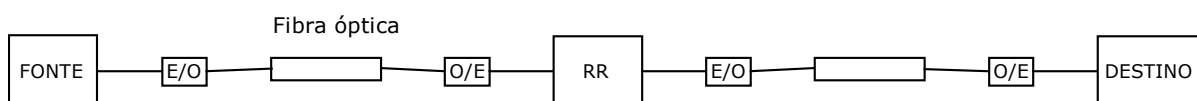



Figura 2-6.5: Meio de transmissão por cabo óptico. O bloco E/O representa o conversor eletroóptico, o bloco O/E representa o conversor optoelétrico e o bloco RR é o Repetidor Regenerador

Atualmente já se usam regeneradores ópticos que não necessitam conversores eletroópticos.

2.6.5. Conversão FDM-TDM

No processo de digitalização da rede, é necessária a interligação dos sistemas FDM e TDM. As tecnologias para conversão FDM-TDM podem ser divididas em: clássica e por transmultiplexação, que se divide em convencional e digital. Ambas tecnologias consistem em demultiplexar o sinal FDM e multiplexá-lo novamente, só que usando TDM.

No sistema europeu, a conversão é realizada de um supergrupo FDM (60 canais) para dois TDM-PCM de 30 canais. No sistema americano, dois grupos FDM são convertidos para um TDM-PCM de 24 canais.

A figura a seguir ilustra a conversão FDM/TDM realizada entre a telefonia urbana e interurbana.

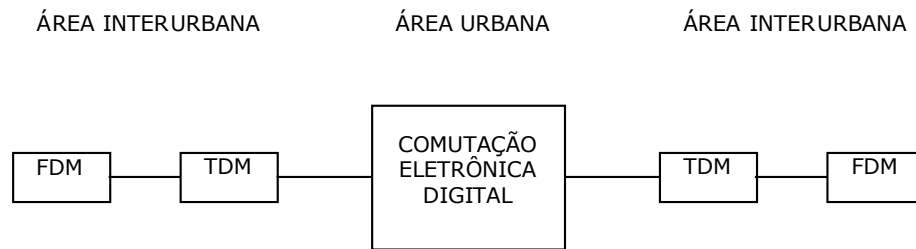



Figura 2-6.6: Conversão FDM/TDM para transmissão

2.7. TRANSFORMAÇÃO DA REDE TELEFÔNICA COM A DIGITALIZAÇÃO

A mudança das centrais telefônicas de analógicas para digitais iniciou-se em 1970. Esta transformação deu origem às centrais digitais CPA-T (Controle por Programa Armazenado - Temporal) e possibilitou a evolução dos métodos de sinalização, passando para a padronização estabelecida pelo SS7, que utiliza um canal dedicado para sinalização.

Em 2002, no Brasil, 98 % das centrais eram digitais, entretanto o enlace do assinante (aparelho telefônico e linha telefônica) ainda é analógico.

2.8. CONFIABILIDADE, DISPONIBILIDADE E SEGURANÇA

A utilização do controle centralizado nas centrais torna a operação dependente de um pequeno número de equipamentos, que devem ser projetados de acordo com elevados padrões de confiabilidade, de forma a apresentarem longos tempos médios entre falhas (*mean time between failure* - MTBF).

Dado que um equipamento falhou, o defeito deve ser encontrado e corrigido no menor tempo possível, o que dá origem a outra medida chamada tempo médio de reparo (*mean time to repair* - MTTR). A disponibilidade de um equipamento é dada pela relação:

$$Disponibilidade = \frac{MTBF}{MTBF + MTTR}$$

A disponibilidade é a probabilidade de o equipamento funcionar corretamente quando requisitado. A probabilidade do equipamento não funcionar é dada por:

$$Indisponibilidade = 1 - Disponibilidade = \frac{MTTR}{MTBF + MTTR}$$

O ITU-T tem recomendações estabelecendo objetivos para a indisponibilidade dos sistemas, listados na tabela 2-8.1.

Abrangência da falha	Objetivo de indisponibilidade
Todo o sistema	$1{,}5 \cdot 10^{-5}$ (6 h em 50 anos)
Linha de assinante	10^{-4} (1 dia em 25 anos)
Circuito entre centrais	10^{-4}
Chamadas de emergência	$1{,}5 \cdot 10^{-5}$
Serviço telefônico básico	10^{-4}
Serviços suplementares	10^{-3}
Tarifação	10^{-4}
Medição de tráfego	10^{-3}
Operações de administração	10^{-2}

Tabela 2-8.1: Objetivos de indisponibilidade

2.9. EVOLUÇÃO DA REDE TELEFÔNICA PARA A RDSI

A existência de redes de comunicação independentes para cada tipo de serviço (voz e dados) não é interessante já que não há compartilhamento dos recursos da rede. A digitalização e codificação dos sinais de comunicação tornaram possível a integração de serviços.

O usuário se beneficia coma a melhora da qualidade, redução de custos e maior variedade de serviços disponíveis. A operadora ganha maior eficiência e uma otimização da rede e dos equipamentos.

Para garantir a interoperabilidade dos serviços, o planejamento e normatização da Rede Integrada devem atender aos seguintes aspectos: versatilidade, velocidade de sinalização, confiabilidade, segurança na transmissão e privacidade na comunicação.

2.10. REDE DIGITAL DE SERVIÇOS INTEGRADOS (RDSI)

2.10.1 Definição

A RDSI (Rede Digital de Serviços Integrados), em inglês: ISDN (*Integrated Services Digital Network*), é uma rede unificada que permite transmissão de voz e dados simultaneamente, proporcionando uma variedade de serviços ao usuário.

A RDSI se caracteriza pelos seguintes aspectos:

- Aproveitamento da atual infra-estrutura do RTPC.
- Integração de Voz, Dados e Imagens na mesma Rede.
- Confiabilidade na Transferência de informações.
- Padronização de interfaces de usuários: a conexão aos diferentes serviços é feita por meio de uma única interface de acesso.
- Transmissão digital entre os terminais.
- Sinalização por um canal separado dos canais de informação: A sinalização é feita através de protocolos de comunicação, dentro de um formato padrão, adotado tanto pelas operadoras de serviços como pelos fabricantes de equipamentos de rede.
- Possibilita comunicação de equipamentos analógicos.

2.10.2. Padronização

O estabelecimento de protocolos padronizados é necessário para garantir a interoperabilidade dos serviços e permitir a ligação de equipamentos de diversos fabricantes à rede. O modelo de referência OSI da ISO deverá ser utilizado para assegurar que a evolução dos protocolos se desenvolva de uma forma estruturada.

2.10.3. Premissas para a formação da RDSI

Há dois níveis de serviços RDSI:

O *acesso básico* (BRI- *Basic Rate Interface*)

- É dedicado aos pequenos usuários e aplicações. Com esta modalidade o usuário pode fazer ligações telefônicas enquanto acessa a internet (a 64 Kbits/s), fazer 2 ligações telefônicas simultâneas, acessar a internet a 128 Kbits/s, passar um fax e usar o telefone ao mesmo tempo, etc.
- O serviço é fornecido pelo mesmo par de fios que chega a sua residência, no entanto, funciona com 3 canais lógicos. Emprega dois canais independentes tipo B de 64 Kbps para transmissão da informação, mais um terceiro canal tipo D para sinalização e controle operando a 16 Kbits/s. Os canais B podem ser combinados para garantir velocidade de acesso de 128 Kbps.
- Um BRI oferece uma taxa líquida de 144 Kbps (2B+D), num circuito a dois fios, utilizando técnicas de comunicação em banda base, modo duplex, por meio de cancelamento de eco.
- A companhia telefônica que oferece o serviço fornece um equipamento chamado NT (*Network terminator*). Todos os equipamentos ISDN devem ser conectados ao NT. Os equipamentos convencionais podem ser conectados a RDSI através de um TA (*Terminal Adapter*), mas a conexão através destes continua sendo analógica.

O *acesso primário* (PRI- *Primary Rate Interface*)

- É destinado a sistemas de maior tráfego. É composto por trinta canais tipo B de 64 Kbps (no padrão europeu), ou 23 canais tipo B (nos Estados Unidos) e um tipo D de 64 Kbps. Nesse caso, a combinação dos trinta canais de transmissão de dados garante uma taxa de até 2 Mbps (E1), numa conexão a quatro fios.
- A linha RDSI é conectada a um PABX digital quando a empresa deseja utilizar a linha para voz ou a um equipamento tipo RAS (Remote Access Server) para a transmissão de dados.
- Aplicações: Videoconferência, Transferência de arquivos, Interligação de Redes Locais de computadores, Acesso remoto à base de dados, Acesso à Internet, etc.

2.10.4. Configurações de Acesso à RDSI

A montagem típica para a interligação de usuários à RDSI é caracterizada pelos seguintes conceitos (Alencar, 1998):

- Ponto de referência: Pontos conceituais usados para separar grupos de funções, com características de interface padronizada. Define as interfaces dos diversos tipos de Serviço.
- Grupamentos funcionais: Arranjos físicos de equipamentos para desempenhar certa função. Descreve as funções dos equipamentos do usuário.

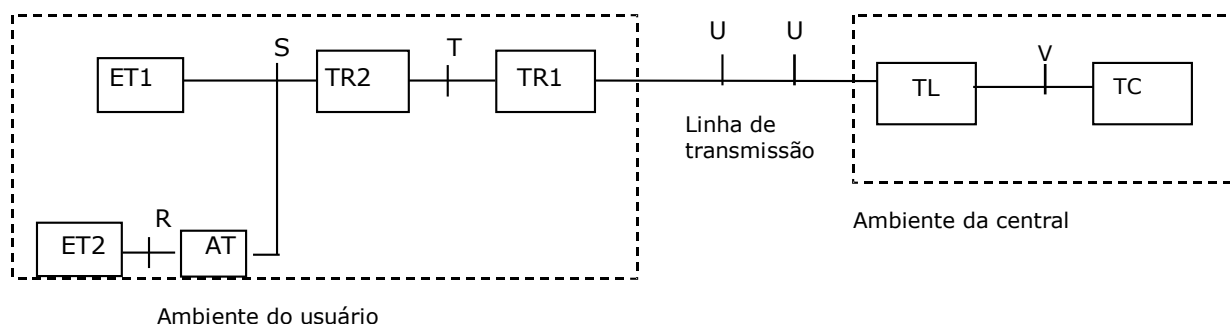



Figura 2-10.1: Configuração típica para RSDI

A figura acima mostra os pontos de referência e grupamentos padronizados, que serão descritos a seguir:

- S e T: Pontos de referência que caracterizam as interfaces físicas entre os equipamentos definidos como tomadas de acesso do usuário à RDSI.
- U: Ponto de referência que pode fazer parte do equipamento de transmissão ou identificar uma interface.
- V: Ponto de referência entre a terminação de linha e a terminação da central.
- ET1: Equipamento terminal RDSI que obedece à interface S da RDSI. Inclui funções de tratamento de protocolo, manutenção, interface e conexão com outros equipamentos. (terminais RDSI puros)
- ET2: Equipamento terminal que inclui as mesmas funções do ET1, mas não obedece à interface S da RDSI. (aparelhos telefônicos comuns, fax, modens analógicos, enfim, qualquer terminal não RDSI).

- AT: Adaptador do terminal, que permite interligar o equipamento ET2 à interface S, ou seja, permite ligar equipamentos analógicos à rede digital.
- TR1: Terminação de rede que faz o acoplamento entre a linha de transmissão e as instalações do usuário. Compreende as funções de terminação de linha, temporização e multiplexação temporal. Faz a conversão do código de linha (interface U) para o código da instalação do usuário (interface interna T). Na linha, a transmissão é feita a dois fios, enquanto na instalação do usuário, pode ser de quatro a oito fios de acordo com o ITU-T.
- TR2: Terminação de rede para distribuição dos terminais ET1 e ET2 nas instalações do usuário, quando este possui instalação multiusuário. O TR2 faz a concentração de acesso de vários terminais. Compreende as funções de tratamento de protocolo, comutação, concentração e manutenção.
- TL: Terminação de linha para acoplamento entre a linha de transmissão da central. Compreende as funções de alimentação, localização de falhas, conversão de códigos e regeneração.
- TC: Terminação de central, que faz o tratamento de sinalização, mantém e supervisiona a conexão do lado da central. Compreende as funções de tratamento de protocolo, gerência de rede, operação e manutenção.

A Central Telefônica

3.1. INTRODUÇÃO

A central telefônica é o elemento de rede responsável pela comutação de sinais entre os usuários. Historicamente, as primeiras centrais desenvolvidas eram sistemas de comutação manual, parte humana parte mecânica, onde a presença da telefonista se fazia imprescindível. Dessa forma, para se completar a chamada, era necessário pedir à telefonista que realizasse a comutação com um ramal determinado. Evidentemente, esse paradigma era deveras enfadonho na medida em que o tempo de comutação estava atrelado a características humanas, impedindo o aumento do número de usuários. Com o passar do tempo, as centrais evoluíram do sistema de comutação manual para o automático, eletromecânico, eletrônico e finalmente digital.

A primeira central pública de programa armazenado, a central 1ESS (*Nº 1 Electronic Switching System*), desenvolvida pela AT&T, foi instalada em *New Jersey*, EUA, em maio de 1965. Este evento histórico deu início ao interesse mundial pela idéia de controle por programa armazenado, que utiliza um processador digital e um programa especialmente desenvolvido para tarefas de comutação (antes designadas à telefonista), gerenciamento e supervisão.

O controle por programa armazenado (*Stored Program Control*) apresenta várias vantagens em relação aos sistemas anteriores.

- **Flexibilidade:** Como a central é controlada por um programa residente que permite alterações é possível, por exemplo, re-configurar a central sem que ela tenha que ser desligada. Isso pode ser feito, inclusive, remotamente pelo fabricante.

- **Facilidades para o assinante:** Centrais de programa armazenado (CPA) permitem um amplo conjunto de facilidades para os assinantes, incluindo:
  - Discagem abreviada;
  - Transferência de chamadas;
  - Restrição às chamadas recebidas;
  - Conta telefônica detalhada;

- Identificação de chamadas maliciosas.

- **Facilidades administrativas:** São facilidades operacionais, do tipo:
    - Controle das facilidades dos assinantes;
    - Mudança no roteamento, para evitar congestionamento de curto prazo;
    - Produção de estatísticas detalhadas do funcionamento da central.

- **Velocidade de estabelecimento da ligação:** As conexões podem ser estabelecidas através de circuitos digitais, reduzindo assim o tempo inicialmente atrelado a características físicas humanas, da ordem de alguns segundos, para o tempo de processamento digital, da ordem de micro segundos. Além disso, a repetição automática das chamadas na própria central pode ser programada, para evitar congestionamento na rede.

- **Economia de espaço:** Ocorre em vista das dimensões reduzidas das centrais de programa armazenado;

- **Facilidade de manutenção:** Os equipamentos da CPA têm uma menor taxa de falhas, em relação aos usados em centrais convencionais, em função de não terem partes móveis;

- **Potencial para outros serviços:** Serviços como transmissão de dados e conferência são facilmente implementados e gerenciados em centrais de programa armazenado;

- **Custo:** As centrais de programa armazenado são mais econômicas para manter e têm um custo de capital mais baixo;

- **Tempo de instalação:** Esse tempo é menor que o necessário para a instalação de centrais analógicas, em virtude da modularização do equipamento digital.

3.2. ESTRUTURA DE UMA CENTRAL TEMPORAL

CLASSIFICAÇÃO de SISTEMAS de COMUNICAÇÃO: Podemos classificar os sistemas de comunicação em relação a vários aspectos, como a seguir. [2]

- **Quanto à conexão**
  - **Sem conexão:** Quando a comunicação ocorre sem a prévia negociação de uma conexão. O sistema postal é um exemplo de comunicação sem conexão, ou seja, as cartas chegam em seu endereço postal sem que você tome conhecimento ou as aceite.
  - **Orientado à conexão (com conexão):** Quando a negociação de uma conexão ocorre antes da comunicação. O sistema telefônico é um exemplo de sistema orientado à conexão na medida em que a ligação precisa ser aceita antes da comunicação começar.

- **Quanto à comutação**
  - **Comutação de circuitos:** Na comutação de circuitos, a reserva de recursos do meio é feita no momento da comutação e perdura até o término da conexão.Dessa forma, é garantida uma quantidade fixa da capacidade do meio para o assinante. Isso significa que um usuário sempre vai possuir aquela quantidade fixa da capacidade, não importa quantos usuários estejam presentes. Se toda capacidade estiver alocada e um usuário adicional tentar fazer-se presente, ele será desprezado.
  - **Comutação de pacotes:** Ao contrário da comutação de circuitos, na comutação de pacotes não existe reserva de recursos do meio, ou seja, a capacidade do canal é compartilhada entre todos os usuários presentes. Isso significa que um usuário pode possuir, em um determinado momento, toda capacidade do canal, assim como pode possuir muito menos que o necessário para realizar o serviço desejado. Na comutação de pacotes nenhum usuário é desprezado por padrão, apesar de existirem mecanismos para tal.

[2] A classificação de sistemas de comunicação será explicada mais detalhadamente em capítulos posteriores.

As centrais telefônicas são interligadas por entroncamentos de fibras ópticas ou cabos de pares em sistemas mais antigos. Elas utilizam comutação de circuitos internamente, o que torna a fase de estabelecimento da ligação a parte mais importante e complexa do processo. As conexões permanecem por toda a duração da chamada. A figura abaixo mostra o diagrama de blocos de uma central hipotética.

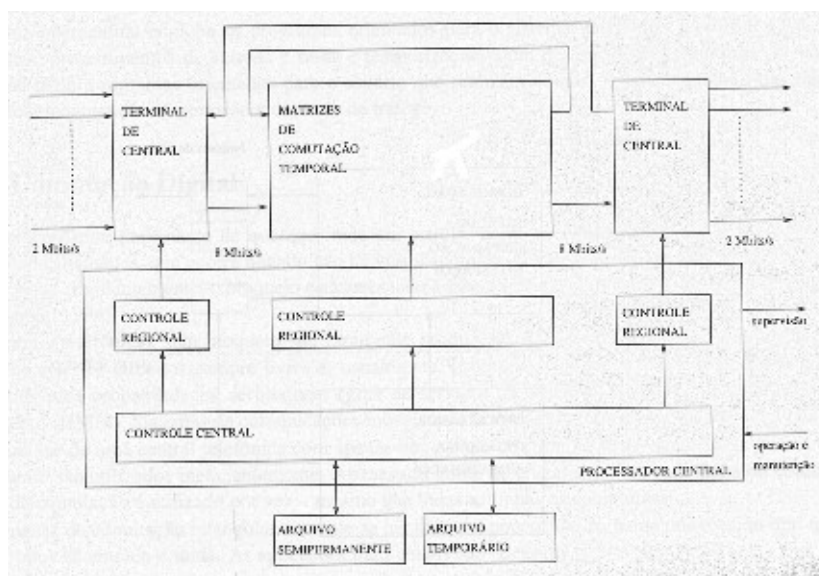



Figura 3-2.1: Diagrama de blocos de um central hipotética

Na figura acima podemos distinguir dois blocos fundamentais para o funcionamento da central: o processador central (onde residem todos os programas de controle, gerenciamento e supervisão) e a matriz de comutação (bloco responsável pela comutação física de circuitos).Nas seções a seguir trataremos desses dois elementos, especificamente.

3.2.1. O processador central

O processador central de uma central CPA-T é geralmente um computador com projeto especializado ou dedicado. O sistema operacional engloba os programas orientados para o sistema, tais como: controle e execução dos programas, procedimentos de entrada e saída e temporização, além de funções de segurança e supervisão. O *software* de aplicação inclui programas orientados para o usuário que realizam tarefas de processamento das ligações, de operação e manutenção, supervisão e controle de tráfego.

3.2.2. A matriz de comutação ou estrutura de comutação

A matriz ou estrutura de comutação é um elemento chave de funcionalidade para a central. Ela realiza a comutação propriamente dita, como o próprio nome indica. Conceitualmente, podemos entender a matriz de comutação como um conjunto entrelaçado de linhas e colunas que se conectam, estabelecendo o caminho físico para a passagem do sinal.

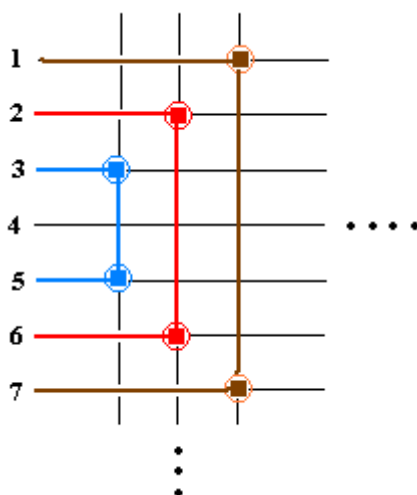



Figura 3-2.2: O conceito de matriz de comutação

Assim, se quiséssemos conectar o ramal número um (disposto na primeira linha) com, digamos o ramal número sete (disposto na sétima linha), utilizaríamos qualquer coluna disponível (no caso a terceira coluna) que funcionaria como ponte condutora para o sinal elétrico, como mostrado na figura. A estrutura de comutação sofreu muitas e radicais mudanças desde sua idealização. As primeiras matrizes era dispositivos manuais (sim, as telefonistas). Com o passar do tempo, elas evoluíram para dispositivos automáticos, eletromecânicos, eletrônicos e finalmente digitais. É importante notar que se todas as colunas estivessem ocupadas, como mostrado na figura, e o assinante de número quatro desejasse realizar uma chamada, ele seria desprezado pelo sistema por falta de recursos, o que significa que o sistema foi sub-dimensionado (o que realmente é feito pelas empresas provedoras de serviços telefônicos). No entanto, para os assinantes que já estão com suas ligações em curso, não importa quantos usuários estejam usando o sistema ou tentando usar, pois eles terão aquele caminho físico disponível mesmo que não estejam se comunicando (silêncio).

3.3. FUNÇÕES DA CENTRAL TELEFÔNICA

As funções principais das centrais telefônicas continuam, basicamente, as mesmas desde a sua invenção.

- **Atendimento:** O sistema monitora constantemente todas as linhas para detectar os pedidos de chamada. O atendimento implica na cessão de recursos para o completamento da chamada;

- **Recepção de informação:** Além dos sinais de solicitação e término da chamada, a central ainda recebe outras informações como endereço da linha chamada e informações relativas a serviços de valor adicionado;

- **Processamento da informação:** O sistema deve processar as informações recebidas para definir as ações a serem tomadas;

- **Teste de ocupado:** Após o processamento da informação e determinação do circuito de saída requerido, o sistema faz um teste de ocupado para verificar a disponibilidade do circuito;

- **Interconexão:** Para uma chamada entre dois usuários, três conexões são realizadas na seqüência seguinte: ligação para o terminal que originou a chamada, ligação com o terminal chamado e conexão entre os dois terminais;

- **Alerta:** Após realizada a conexão, o sistema alerta o usuário chamado e envia um tom característico para o assinante que chama;

- **Supervisão:** A supervisão da chamada é feita durante todo o tempo para tarifação e determinação do instante em que o circuito deve ser desconectado;

- **Envio de informação:** Ocorre sempre que o usuário se encontre ligado a outra central. A central de origem deve enviar, por exemplo, a informação de endereço para ser processada pela central de destino.

As funções referentes ao estabelecimento de conexões são de extrema importância. Dessa forma, o estabelecimento da conexão é um processo que merece destaque e será abordado com mais detalhe na próxima seção.

3.4. ESTABELECIMENTO DE UMA CONEXÃO

Nessa seção abordaremos o procedimento executado por uma central para o estabelecimento de uma ligação telefônica [3]. Muitas coisas podem acontecer quando tentamos realizar uma ligação, no entanto ilustraremos aqui o caso em que a chamada é completada com sucesso. Outros casos possíveis seriam: ramal ocupado, número inexistente e etc...

Suponha que Bob queira conversar com Alice. Para saciar o seu desejo, nosso amigo Bob vai até o seu telefone e tira o fone do gancho. Nesse momento, a central telefônica diretamente conectada ao ramal do Bob identifica a mudança de estado (retirada do gancho do ramal do Bob) e retorna ao nosso amigo o tom de discar, caso haja recursos disponíveis no sistema para completar a chamada do Bob (neste caso assumiremos que a chamada será completada). A partir desse momento, o usuário, no caso o Bob, sabe que tem permissão para discar o número do telefone de sua amiga Alice. Ao pressionar a primeira tecla do número de Alice, a central interrompe o tom de discar para notificar ao usuário que está identificando os dígitos pressionados. Após realizar toda discagem, e supondo que Alice não está utilizando seu telefone nesse momento, a central telefônica retorna ao nosso querido amigo o tom de chamando, enquanto envia ao ramal de Alice o *ring* (tom de chamando para o destinatário). Nesse momento, Alice percebe que alguém deseja falar com ela e retira o seu fone do gancho para iniciar a conversação (momento de estabelecimento ou aceitação da conexão). Ao perceber que Alice retirou o fone do gancho, a central interrompe o sinal de *ring* e conecta finalmente os dois usuários para que eles possam conversar em paz. Quando terminarem de falar, Bob irá colocar o seu fone novamente no gancho, notificando à central que terminou sua ligação, e a central terminará a conexão entre os dois além de liberar os recursos alocados por Bob no início do telefonema. O esquema explicado anteriormente está ilustrado na figura abaixo, onde Bob é indicado como usuário A, realizador da chamada, e Alice como usuário B, receptor da chamada.

---

[3] Toda a sinalização utilizada pela central para o estabelecimento da ligação telefônica está explicada no capítulo 2

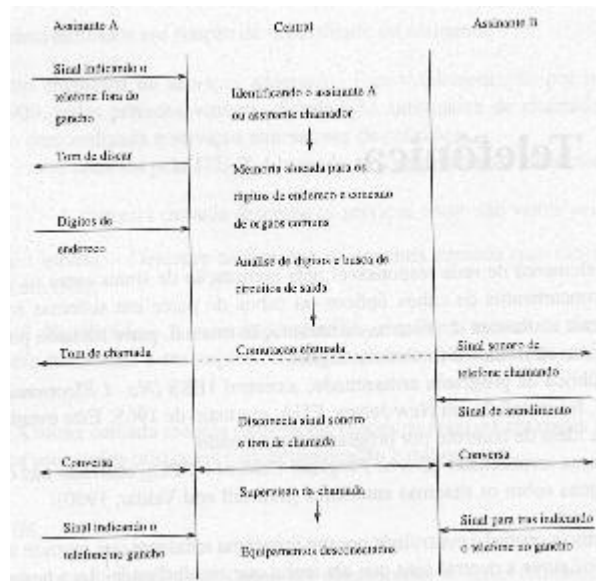



Figura 3-4.1: Diagrama de eventos de uma ligação completada

3.5. A FAMÍLIA TRÓPICO

O sistema Trópico é uma família completa de centrais digitais de comutação telefônica, alcançando de 192 a mais de 50.000 terminais disponíveis em duas versões. A primeira versão, o trópico R, é de pequena/média capacidade enquanto que a segunda Trópico RA é de média/alta capacidade.

A Trópico foi construída segundo uma concepção de módulos independentes. Cada módulo tem seu próprio microprocessador e seu software. Em funcionamento normal, os módulos trocam informações entre si, podem ser gerenciados à distância e os seus dados modificados pelo operador. Em caso de falhas, a degradação ocorre de forma suave, pois só o módulo com defeito pára de funcionar.

A família Trópico compreende um concentrador de linhas de assinantes, em produção desde 1983 (Trópico C). A central local de pequeno porte para 4.000 assinantes e 800 troncos está no mercado desde 1985 (Trópico R). A central local/tandem de médio porte para 16 mil linhas iniciou a produção em 1991 (Trópico RA). Ainda em desenvolvimento estão as centrais Trópico L, local/tandem de grande porte para 80.000 linhas e a Trópico T, interurbana de grande porte para 50.000 circuitos troncos.

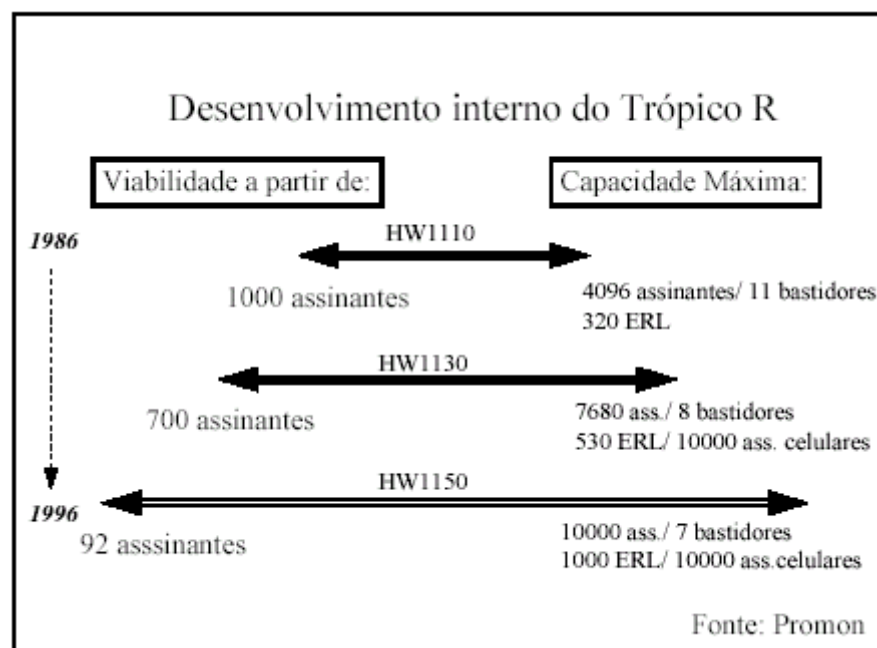



Figura 3-5.1: A evolução das centrais Trópico R

3.6. CENTRAIS TRÓPICO R

A central Trópico R tem sua estrutura dividida em duas partes: *hardware* e *software*. A estrutura *hardware* por sua vez pode ser subdividida em:

- **Estrutura física;**
- **Estrutura funcional.**

3.6.1. Estrutura Física

As centrais Trópico são constituídas de três módulos: módulo de terminais (MT); módulo de comutação (MC) e módulo de operação e manutenção (MO). Estes módulos são implementados em um sistema mecânico horizontal padrão composto de bastidores, sub-bastidores, elementos e placas de circuito impresso.

- **Bastidores**: São armários numerados, onde a numeração varia dependendo da capacidade, esses armários são divididos em sub-bastidores;

- **Sub-bastidores**: São as gavetas onde serão inseridas as placas de circuito impresso. Cada sub-bastidor acomoda um ou mais submódulos da central e tem sempre uma fonte de alimentação para o fornecimento das tensões e correntes elétricas requeridas pelas placas nele inseridas;

- **Elementos**: São considerados elementos as unidades que possuem características diferentes dos sub-bastidores, são eles:
    - **Elementos de distribuição de energia e gerador de toque (EDEG**);
    - **Elemento módulo de operação portátil**;
    - **Elemento de acesso a periféricos seriais**.

3.6.2. Estrutura Funcional

Corresponde à estrutura que define os módulos que suprem as operações que garantem o funcionamento de uma central telefônica.

Módulo é um elemento funcional, implementado fisicamente em um sub-bastidor, caracterizado pelo fato de conter até quatro submódulos e uma interface de acesso aos planos (IAP-T) de comutação da central. O submódulo é o elemento funcional que possui um controlador (processador) podendo ter sua função (hardware) distribuída por uma ou mais placas para atender a funções específicas. O número de planos de comutação em um central Trópico R varia de um a no máximo três. Os três tipos de módulos existentes que já foram citados são os seguintes:

- **Módulos de Terminais (MT);**
- **Módulo de Comutação (MC);**
- **Módulo de Operação e Manutenção (MO)**.

A evolução das centrais, através da compactação e do empacotamento realizado, possibilitou a existência de um único módulo de comutação e operação (MCO), substituindo os módulos de comutação e módulo de operação e manutenção que se apresentavam em módulos separados.

3.7. ESTRUTURA DO SOFTWARE DA CENTRAL

O software das centrais Trópico R possui uma estrutura que agrupa todas as funções pertinentes a uma central telefônica, associada às funções pertinentes à tecnologia digital. Esses agrupamentos que cumprem funções inerentes ao processamento, supervisão e gerência de um sistema telefônico, deram origem ao que se denomina Blocos de Implementação (BI). Um BI *software* é a menor divisão da estrutura funcional do *software* de uma central Trópico R. Assim, um BI é um programa que tem por objetivo administrar um dado recurso da central. Estes BIs residem nas memórias dos processadores (CTS/CVS/CMG/CTF), também denominados de controladores e que estão espalhados por toda a estrutura física da central.

As funções realizadas pela central estão separadas em funções básicas e de aplicação. As funções básicas são aquelas destinadas a dotar o sistema de uma infra-estrutura capaz de permitir o funcionamento do sistema, por exemplo, controle de processadores, calendário e alarmes.

As funções de aplicação são as que se utilizam de estrutura básica para controlar e supervisionar os recursos da central tais como: terminais de assinantes, juntores, registradores, tarifação e comunicação homem-máquina. Os BIs software assim como as funções da central também foram divididos em BIs de sistema básico e BIs de sistema de aplicação.

3.7.1. Sistema básico

Constitui-se do sistema operacional dos diversos processadores, controlando toda a comunicação entre esses processadores, além de fornecer suporte às funções telefônicas. Em todos os processadores da central existem obrigatoriamente o núcleo operacional, constituído pelos BIs NUI e NUC além dos BIs do sistema operacional formados pelos BIs BPR e BRF. Os demais BIs controlam as diversas funções básicas inerentes às funções telefônicas.

3.7.2. Sistema de Aplicação

Constituído dos BIs controladores e supervisores de recursos telefônicos, responsáveis pelo tratamento de chamadas telefônicas e pela operação e manutenção dos recursos da central.

Para tratamento de chamadas, os BIs estão associados ou a terminais físicos ou a tabelas de dados nas quais está implementada a lógica de análise de dígitos e encaminhamento da central. Para operação e manutenção, os BIs centralizam a gerência de dados relativos aos recursos da central. Servem como interface entre o operador e os controladores de equipamento da central permitindo a realização de funções, tais como criar, suprimir, bloquear, ativar e alterar dados.

3.8. ESTRUTURAS DE GERENCIAMENTO E SUPERVISÃO

3.8.1 Estrutura de Dados

Nas centrais Trópico R, a estrutura de dados possui a característica de ser alterada com pouca freqüência. Tais dados, denominados semipermanentes, são armazenados na central, em disquetes e nas memórias das placas de processadores. As informações que compõem esses dados semipermanentes podem ser alteradas via monitor da central ou via disquete.

Dados alterados via monitor:

- **Criação e suspensão de assinantes, juntores;**
- **Categoria e restrições de tráfego de assinantes;**
- **Informações de calendários, feriados e outras;**
- **Bloqueio e desbloqueio de processadores, por exemplo.**

Dados alterados via disquete:

- **Dados de tarifação;**
- **Dados de rota;**
- **Tabela para análise de encaminhamento**.

A base de dados da central possui três níveis de armazenamento que são:

- **Primário**: Constitui-se de elementos localizados nos controladores de recursos, que detêm dados relacionados ao submódulo controlado;
- **Central**: Formado pelos controladores centralizados nos supervisores de recurso;
- **Memória de massa**: Constituída de até duas unidades de disco magnético flexível 3 ½ que armazenam as tabelas de dados semipermanentes e de tarifação, além dos registros de falhas da central.

3.8.2. Tarifação

Os pulsos de tarifação, durante cada chamada, são armazenados em contadores individuais existentes nos BIs controladores de assinantes (memória RAM).

Ao término de cada chamada, os dados de tarifação correspondentes são copiados para contadores centralizados (BI ATX- supervisor de tarifação) sendo armazenados em memória de massa a cada duas horas.

3.8.3. Supervisão e falhas

As falhas ocorridas no sistema são detectadas a nível de processador pelo BI BRF (controlador de registro de falhas), que é responsável pela supervisão de falhas de todos os BIs de seu respectivo processador. Estes alarmes são sintetizados e enviados ao BI BAL (controlador de alarmes), responsável pela gerência de todos os alarmes da central. Este analisa as falhas ocorridas e, de acordo com a categoria, envia um resumo das falhas aos BIs ATA (controlador de terminal alfanumérico) a ATN (controlador de terminal numérico) para sinalização ao operador.

3.9. CARACTERÍTICAS DA CENTRAL TRÓPICO RA

A central trópico RA é uma central digital com tecnologia CPA-t, 100% nacional, desenvolvida no CPqP da Telebrás, em parceria com três fabricantes: Promon, STc e Alcatel.

As principais características dessa central são:

- **Comutação temporal com conversão de sinais analógico/digital e digital/analógico ao nível de terminais;**
- **Controle distribuído e descentralizado;**
- **Participação de carga sem troca de mensagens de atualização entre processadores;**
- **Degradação suave em caso de falhas;**
- **Estrutura de voz, sinalização e sincronismo independentes;**
- **Redundância ativa nas funções de comutação, sinalização e distribuição de sincronismo;**
- **Sincronismo mestre/escravo;**
- **Padronizações de interfaces;**
- **Alto grau de modularidade e expansão;**
- **Alta capacidade para absorção de evoluções tecnológicas;**
- **Bilhetagem.**

A central possui capacidade para 32.000 assinantes, 3.600 Erl de tráfego comutado, 680.000 chamadas por hora, 5.460 rotas, 31.200 juntores e 1.024 processadores. Em termos de encaminhamento de chamada há cinco planos de encaminhamento, marcação de origem, interceptação automática, encaminhamento para máquina anunciadora centralizada e conexão semipermanente.

Os seguintes serviços suplementares são oferecidos: Discagem abreviada, Linha direta (hot line), Linha executiva, Restrição de chamadas originadas, Controle de restrição pelo assinante, Transferência automática em caso de não responde, Não perturbe, Prioridade, Registro detalhado de chamadas originadas, Chamadas registradas, Despertador automático, chamada em espera, Consulta, Conferência e Identificação do assinante chamador (BINA).

A central Trópico RA permite interface com telefones decádicos e multifreqüênciais, telefone público, acesso digital a 64 Kbit/s, CPCT com ou sem DDR, central telefônica

comunitária (CTC), concentradores e respondedores automáticos. Os entroncamentos são efetivados através de juntores digitais, analógicos a dois e quatro fios, juntores para mesas operadoras e juntores para máquina anunciadora.

A tarifação é efetuada por meio de bilhetagem automática, multimedição e serviço medido local com 16 grupos de modalidades de tarifa agendados independentemente, 256 classes de cadência e 128 códigos de tarifa reversa para assinante e rota de saída e transmissão de dados de tarifação remota.

Além da sinalização MFC, as centrais Trópico RA, utilizam a sinalização por canal comum. A sinalização comum é composta por: sub-sistema de transferência de mensagens, sub-sistema de usuário RDSI, sub-sistema de usuário telefônico e sistema de controle de conexão de sinalização.

As principais características operacionais são:

- **Condições para criar, suprimir, modificar e interrogar dados de assinantes, juntores, encaminhamentos, sinalização, tarifação e configuração da central**;
- **Medição de tráfego com resultados dos últimos 15 minutos, da hora de medição, da HMM (hora de maior movimento) da central e de grupos de órgãos das últimas 24 horas;**
- **Emissão de relatórios por órgãos e grupo de órgãos;**
- **Emissão de relatórios de supervisão de tráfego, de densidade de chamadas e de tráfego por assinantes;**
- **Relatório configurável para registro de dados de até 2.000 chamadas.**

Em termos de manutenção as principais características são:

- **Detecção e localização de falhas de processadores, placas de terminais e fontes de alimentação;**
- **Sinalização espontânea de falhas de processadores, placas de terminais e fontes de alimentação;**
- **Chamada de teste dirigida;**
- **Geração de chamada no circuito de linha do assinante**.

As principais características de supervisão são:

- **Supervisão de desempenho dos órgãos e grupos de órgãos da central;**
- **Relatório de desempenho, exceção e resumo de exceção;**
- **Bloqueio automático dos órgãos com quantidades sucessivas de ocupações ineficientes;**
- **Supervisão da qualidade de transmissão;**
- **Supervisão de temporizações;**
- **Supervisão de telefone público;**
- **Supervisão de chamada maliciosa;**
- **Registro de chamadas para tarifação.**

Comunicações Ópticas

4.1 INTRODUÇÃO

A idéia de transmitir informações por meio de sinais luminosos não é recente. Relata-se como um dos feitos mais notáveis, a primeira transmissão de voz feita, em 1880, por Graham Bell por meio de um feixe luminoso.

A fibra ótica, nada mais é do que uma tecnologia na qual a luz é transmitida ao longo da parte interna e flexível da fibra de vidro ou plástica, que forma o núcleo das fibras. Seu desenvolvimento teve uma evolução muito rápida nos últimos 30 anos. Desde a década de 60, a atenuação foi reduzida de várias ordens de grandeza, a capacidade de transmissão aumentada enormemente e as aplicações difundidas pelos mais diversos campos de utilização.

4.2. ALGUMAS DEFINIÇÕES IMPORTANTES

4.2.1. A luz como fenômeno

Duas teorias explicam atualmente a natureza da luz: teoria corpuscular e a teoria ondulatória. A teoria ondulatória, define a luz como campos eletromagnéticos propagando-se no espaço. Explica fenômenos como reflexão, refração, difração etc. Já a teoria corpuscular, define a luz como pacotes de energia denominados fótons e explica fenômenos como efeito Compton e o desvio de raios luminosos que passem perto de corpos celestes.

4.2.2. Índice de Refração

É a grandeza que expressa a razão entre a velocidade da luz no vácuo e a velocidade da luz no meio em que ela se propaga. É definida por $n = c/v$, onde

- c é a velocidade da luz no vácuo
- v é a velocidade da luz no meio em questão

Nas fibras ópticas o índice de refração poderá provocar dispersão do impulso luminoso, limitando a capacidade de transmissão. Este fenômeno explica a decomposição da luz branca no prisma e a formação do arco-íris.

Ex: vácuo n=1,0; água n=1,3; vidro n=1,5; diamantes n=2,0.

4.2.3. Reflexão e Refração

Ao incidir em uma superfície de separação entre dois meios, de índice de refração diferentes, uma parcela da energia será refletida enquanto outra parcela se propagará através do segundo meio. Ao passar para o meio de transmissão, a onda sofre um desvio na sua direção caracterizado pela Lei de Snell:

$$n_1 \cdot sen\,\theta_1 = n_2 \cdot sen\,\theta_2$$

4.2.4. Ângulo crítico e reflexão interna total

Quando um raio de luz muda de um meio que tem índice de refração maior para um meio que tem índice de refração menor, a direção da onda transmitida afasta-se da normal. A medida que aumentamos o ângulo de incidência i, o ângulo do raio refratado tende a 90º. Quando isso acontece, o ângulo de incidência recebe o nome de ângulo crítico. Uma incidência com ângulo maior do que este sofre o fenômeno da reflexão interna total.

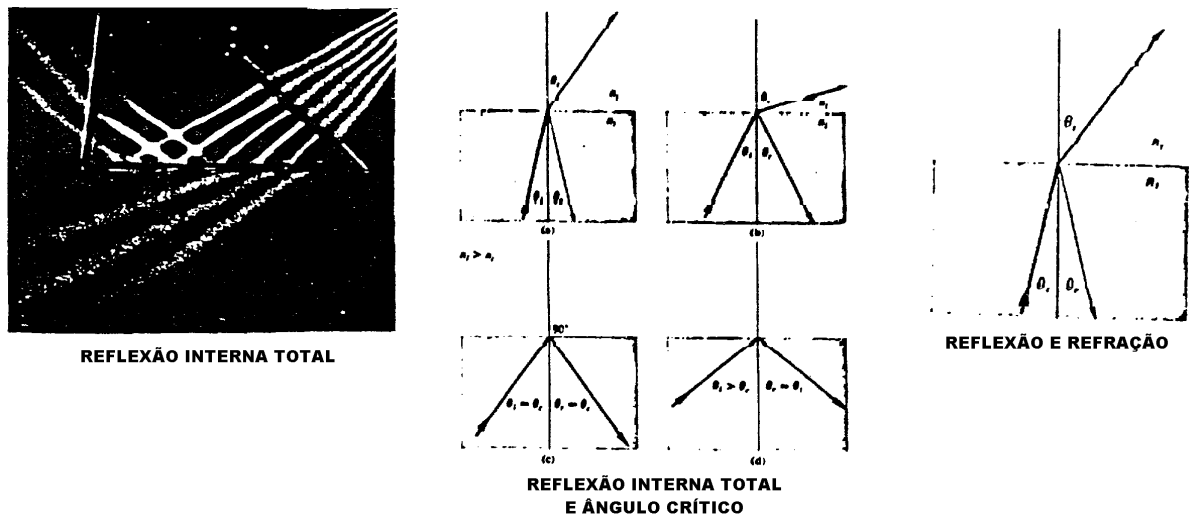



Figura 4-2.1: Reflexão e refração

4.3. FIBRAS ÓPTICAS

Uma fibra óptica é um capilar formado por materiais cristalinos e homogêneos, transparentes o bastante para guiar um feixe de luz (visível ou infravermelho) através de um trajeto qualquer. A estrutura básica desses capilares são cilindros concêntricos com determinadas espessuras e com índices de refração tais que permitam o fenômeno da reflexão interna total. O centro (miolo) da fibra é chamado de núcleo e a região externa é chamada de casca. Para que ocorra o fenômeno da reflexão interna total é necessário que o índice de refração do núcleo seja maior que o índice de refração da casca. Os tipos básicos de fibras ópticas são:

- fibra de índice degrau
- fibra de índice gradual
- fibra monomodo

4.3.1. Fibra de Índice Degrau (Step Index)

Este tipo de fibra foi o primeiro a surgir e é o tipo mais simples. Constitui-se basicamente de um único tipo de vidro para compor o núcleo, ou seja, com índice de refração constante. O núcleo pode ser feito de vários materiais como plástico, vidro, etc. e com dimensões que variam de 50 a 400 μ m, conforme o tipo de aplicação.

A casca, cuja a função básica é garantir a condição de guiamento da luz pode ser feita de vidro ou plástico e até mesmo o próprio ar pode atuar como casca.

Essas últimas, chamadas de *bundle,* são limitadas quanto à capacidade de transmissão; possuem atenuação elevada (maior que 5 dB/km) e pequena largura de banda (menor que 30 MHz.km) e são utilizadas em transmissão de dados em curtas distâncias e iluminação.

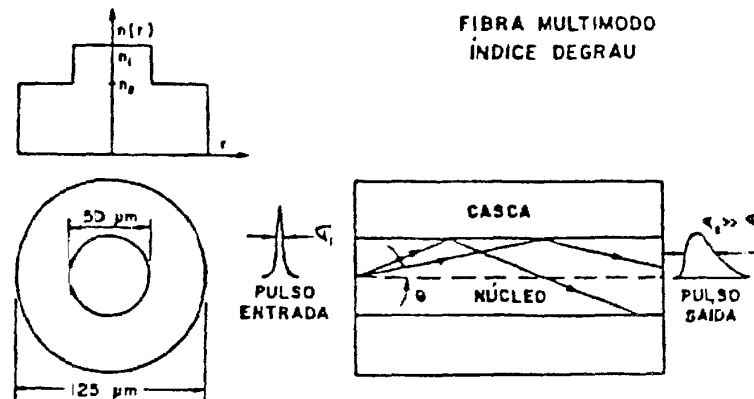



Figura 4-3.1: Fibra multímodo índice degrau

4.3.2. Fibra de Índice Gradual (Graded Index)

Este tipo de fibra tem seu núcleo composto por vidros especiais com diferentes valores de índice de refração, os quais têm o objetivo de diminuir as diferenças de tempos de propagação da luz no núcleo, devido aos vários caminhos possíveis que a luz pode tomar no interior da fibra, diminuindo a dispersão do impulso e aumentando a largura de banda passante da fibra óptica.

A variação do índice de refração em função do raio do núcleo obedece à seguinte equação $n(r) = n_1 \cdot (1-(r/a) \cdot \alpha \cdot \Delta)$, onde

- n(r) é o índice de refração do ponto r
- n_1 é o índice de refração do núcleo
- r é a posição sobre o raio do núcleo
- α é o coeficiente de otimização
- Δ é a diferença entre o índice de refração da casca e do núcleo

Os materiais tipicamente empregados na fabricação dessas fibras são sílica pura para a casca e sílica dopada para o núcleo com dimensões típicas de 125 e 50 μm respectivamente. Essas fibras apresentam baixas atenuações (3 db/km em 850 nm) e capacidade de transmissão elevadas. São, por esse motivo, empregadas em telecomunicações.

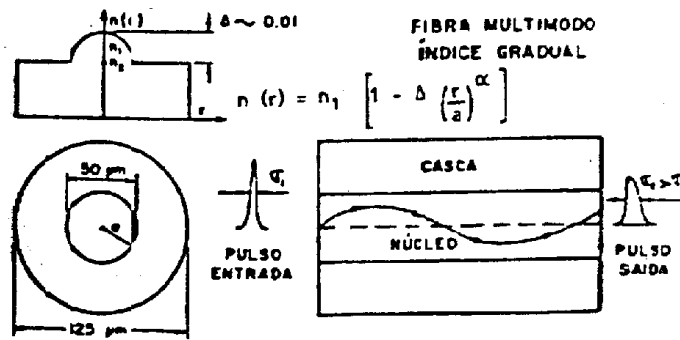



Figura 4-3.2: Fibra multimodo índice gradual

4.3.3. Fibra Monomodo

Esta fibra, ao contrário das anteriores, é construída de tal forma que apenas o modo fundamental de distribuição eletromagnética (raio axial) é guiado, evitando assim os vários caminhos de propagação da luz dentro do núcleo, conseqüentemente diminuindo a dispersão do impulso luminoso.

Para que isso ocorra, é necessário que o diâmetro do núcleo seja poucas vezes maior que o comprimento de onda da luz utilizado para a transmissão. As dimensões típicas são 2 a 10 μm para o núcleo e 80 a 125 μm para a casca. Os materiais utilizados para a sua fabricação são sílica e sílica dopada.

São empregadas basicamente em telecomunicações pois possuem baixa atenuação (0,7 dB/km em 1300 nm e 0,2 dB/km em 1550 nm) e grande largura de banda (10 a 100 GHz.km).

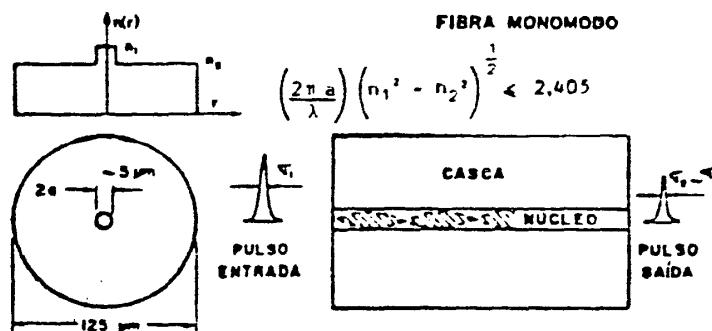



Figura 4-3.3: Fibra monomodo

4.4. GUIAMENTO DE LUZ EM FIBRAS ÓPTICAS

4.4.1. Abertura Numérica

É um parâmetro básico para fibras multimodos (degrau e gradual) que representa o ângulo máximo de incidência que um raio deve ter, em relação ao eixo da fibra, para que ele sofra a reflexão interna total no interior do núcleo e propague-se ao longo da fibra através de reflexões sucessivas.

4.4.2. Modos de Propagação

Quando tratamos a luz pela teoria ondulatória, a luz é regida pelas equações de Maxwell. Assim, se resolvermos as equações de Maxwell para as condições (chamadas condições de contorno) da fibra, que é um guia de onda, tais como diâmetro do núcleo, comprimento de onda, abertura numérica, etc. encontramos um certo número finito de soluções. Dessa maneira, a luz que percorre a fibra óptica não se propaga aleatoriamente, mas é canalizada em certos modos.

Modo de propagação é, portanto, uma onda com determinada distribuição de campo eletromagnético que satisfaz as equações de Maxwell e que transporta uma parcela individual (mas não igual) da energia luminosa total transmitida. Esses modos podem ser entendidos e representados como sendo os possíveis caminhos que a luz pode ter no interior do núcleo.

4.5. PROPRIEDADES DAS FIBRAS ÓPTICAS

A fibra óptica apresenta certas características particulares, que podemos tratar como vantagens, quando comparadas com os meios de transmissão formados por condutores metálicos, tais como:

- Imunidade a ruídos externos em geral e interferências eletromagnéticas em particular, como as causadas por descargas atmosféricas e instalações elétricas de altas tensões;
- Imunidade a interferências de freqüências de rádio de estações de rádio e radar, e impulsos eletromagnéticos causados por explosões nucleares;
- Imunidade a influência do meio ambiente, como por exemplo umidade;
- Ausência de diafonia;
- Grande confiabilidade no que diz respeito ao sigilo das informações transmitidas;
- Capacidade de transmissão muito superior a dos meios que utilizam condutores metálicos;
- Baixa atenuação, grandes distâncias entre pontos de regeneração;
- Cabos de pequenas dimensões (pequeno diâmetro e pequeno peso) o que implica em economia no transporte e instalação.

4.6. APLICAÇÕES DAS FIBRAS ÓPTICAS

- Redes de telecomunicações
  - Entroncamentos locais
  - Entroncamentos interurbanos
  - Conexões de assinantes
- Redes de comunicação em ferrovias
- Redes de distribuição de energia elétrica (monitoração, controle e proteção)
- Redes de transmissão de dados e fac-símile
- Redes de distribuição de radiodifusão e televisão
- Redes de estúdios, cabos de câmeras de TV
- Redes internas industriais
- Equipamentos de sistemas militares
- Aplicações de controle em geral
- Veículos motorizados, aeronaves, navios, instrumentos, etc.

4.7. CARACTERÍSTICAS DE TRANSMISSÃO

4.7.1. Atenuação

A atenuação ou perda de transmissão pode ser definida como a diminuição da intensidade de energia de um sinal ao propagar-se através de um meio de transmissão. A fórmula mais usual para o cálculo da atenuação em Decibéis é a seguinte $10 \cdot \log \frac{P_s}{P_e}$, onde

- P_s é a potência de saída
- P_e é a potência de entrada

Nas fibras ópticas, a atenuação varia de acordo com o comprimento de onda da luz utilizada. Essa atenuação é a soma de várias perdas ligadas ao material que é empregado na fabricação das fibras e à estrutura do guia de onda. Os mecanismos que provocam atenuação são:

- absorção
- espalhamento
- deformações mecânicas

4.7.2. Absorção

Os tipos básicos são a absorção material e a absorção do íon OH^-.

A absorção material é o mecanismo de atenuação que exprime a dissipação de parte da energia transmitida numa fibra óptica em forma de calor. Neste tipo de absorção temos fatores extrínsecos e intrínsecos à própria fibra.

Como fatores intrínsecos, temos a absorção do ultravioleta, a qual cresce exponencialmente no sentido do ultravioleta, e a absorção do infravermelho, provocada pela

vibração e rotação dos átomos em torno da sua posição de equilíbrio, a qual cresce exponencialmente no sentido do infravermelho.

Como fatores extrínsecos, temos a absorção devida aos íons metálicos porventura presentes na fibra (Mn, Ni, Cr, U, Co, Fe e Cu) os quais, devido ao seu tamanho, provocam picos de absorção em determinados comprimentos de onda exigindo grande purificação dos materiais que compõem a estrutura da fibra óptica.

4.7.3. Espalhamento

É o mecanismo de atenuação que exprime o desvio de parte da energia luminosa guiada pelos vários modos de propagação em várias direções. Existem vários tipos de espalhamento (Rayleigh, Mie, Raman estimulado, Brillouin estimulado) sendo o mais importante e significativo o espalhamento de Rayleigh. Esse espalhamento é devido à não homogeneidade microscópica (de flutuações térmicas, flutuações de composição, variação de pressão, pequenas bolhas, variação no perfil de índice de refração, etc).

Esse espalhamento está sempre presente na fibra óptica e determina o limite mínimo de atenuação nas fibras de sílica na região de baixa atenuação. A atenuação neste tipo de espalhamento é proporcional a $\frac{1}{\lambda^4}$.

4.7.4. Deformações Mecânicas

As deformações, chamadas de *microcurvatura* e *macrocurvatura*, ocorrem ao longo da fibra devidas à aplicação de esforços sobre a mesma durante a confecção e instalação do cabo.

As macrocurvaturas são perdas pontuais (localizadas) de luz por irradiação, ou seja, os modos de alta ordem (ângulo de incidência próximo ao ângulo crítico) não apresentam condições de reflexão interna total devidas às curvaturas de raio finito da fibra óptica.

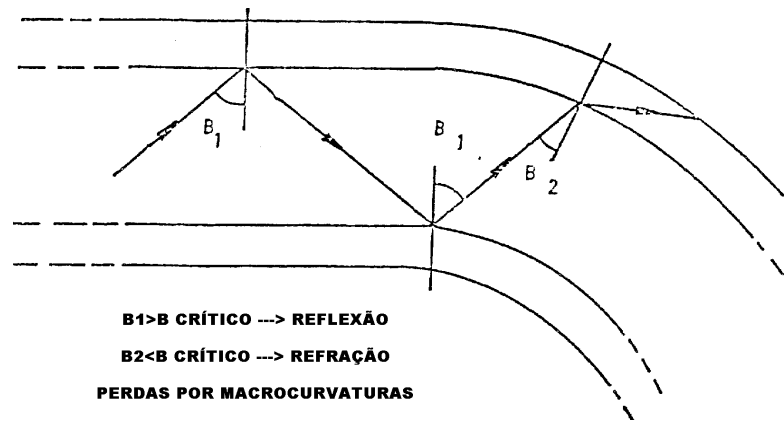



Figura 4-7.1: Deformações Mecânicas

As microcurvaturas aparecem quando a fibra é submetida à pressão transversal de maneira a comprimi-la contra uma superfície levemente rugosa. Essas microcurvaturas extraem parte da energia luminosa do núcleo devido aos modos de alta ordem tornarem-se não guiados.

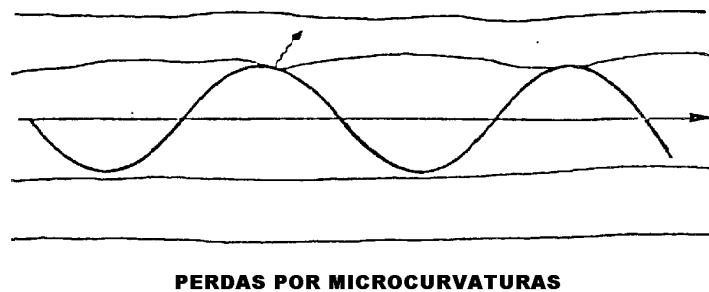



Figura 4-7.2: Perdas por microcurvaturas

A atenuação típica de uma fibra de sílica sobrepondo-se todos os efeitos está mostrada na figura abaixo:

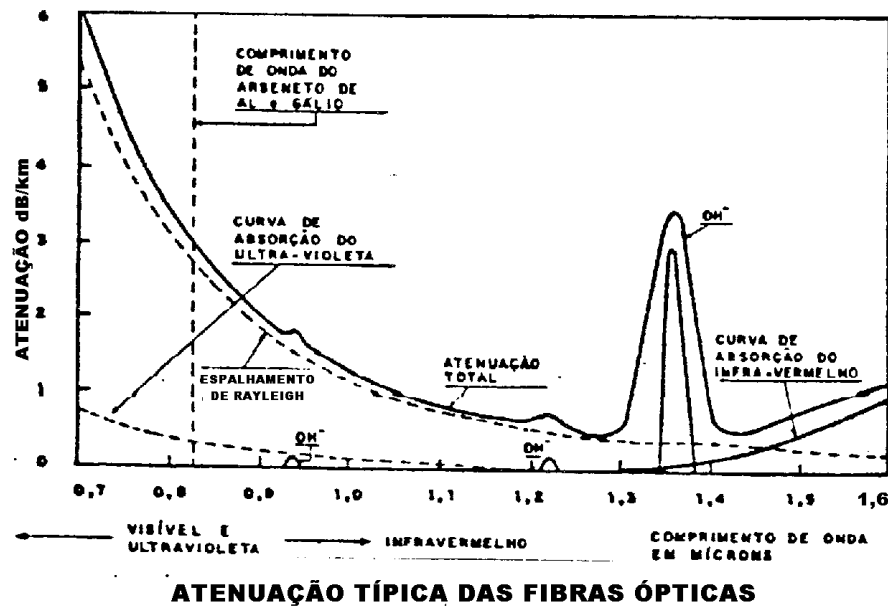



Figura 4-7.3: Atenuação típica das fibras ópticas

Existem três λ 's tipicamente utilizados para transmissão em fibras ópticas:

- 850 nm com atenuação típica de 3 dB/km
- 1300 nm com atenuação típica de 0,8 dB/km
- 1550 nm com atenuação típica de 0,2 dB/km

4.7.5. Dispersão

É uma característica de transmissão que exprime o alargamento dos pulsos transmitidos. Este alargamento determina a largura de banda da fibra óptica, dada em MHz.km, e está relacionada com a capacidade de transmissão de informação das fibras. Os mecanismos básicos de dispersão são modal, existente em fibras multimodo, provocada pelos caminhos possíveis de propagação (modos) que a luz pode ter no núcleo e cromática, que depende do comprimento de onda e divide-se em dois tipos: Material e de Guia de Onda. Os tipos de dispersão que predominam nas fibras são:

- Degrau: Modal (dezenas de MHz.Km).
- Gradual: Modal Material (menor que 1 GHz.Km).

- Monomodo: Material e Guia de Onda (10 a 100 GHz.Km).

4.8. MÉTODOS DE FABRICAÇÃO

Os materiais básicos usados na fabricação de fibras ópticas são sílica pura ou dopada, vidro composto e plástico. Todos os processos de fabricação são complexos e caros.

As fibras ópticas fabricadas com sílica pura ou dopada são as que apresentam as melhores características de transmissão e são as usadas em sistemas de telecomunicações. As fibras ópticas fabricadas com vidro composto e plástico não têm boas características de transmissão (possuem alta atenuação e baixa largura de banda passante) e são empregadas em sistemas de telecomunicações de baixa capacidade e pequenas distâncias e sistemas de iluminação. Os processos de fabricação dessas fibras são simples e baratos se comparados com as fibras de sílica pura ou dopada.

4.9. CABOS ÓPTICOS

O uso de fibras ópticas gerou uma série de modificações nos conceitos de projeto e fabricação de cabos para telecomunicações. Nos cabos de condutores metálicos as propriedades de transmissão eram definidas pelo condutor, construção do cabo e materiais isolantes. Estes cabos eram pouco afetados nas suas características pelas trações e torções exercidas sobre os cabos durante a fabricação e instalação. Já nos cabos ópticos, a situação é diferente porque as características de transmissão dependem apenas da fibra óptica e sua fragilidade é notória. No projeto de cabos ópticos são observados os seguintes ítens:

- Número de fibras
- Aplicação
- Minimização de atenuação por curvaturas
- Características de transmissão estável dentro da maior gama de temperaturas possível
- Resistência à tração, curvatura, vibração, compressão adequadas
- Degradação com o tempo (envelhecimento)
- Facilidade de manuseio, instalação, confecção de emendas, etc.

4.10. MEDIDAS EM FIBRAS ÓPTICAS

Para a caracterização das fibras ópticas são efetuadas medições que verificam as características de transmissão das fibras, a saber:

- Atenuação espectral
- Atenuação de inserção
- Largura de banda
- Abertura numérica
- Perfil de índice de refração

4.10.1. Test e de Atenuação Espectral

Este tipo de teste mede a atenuação da fibra óptica numa faixa de comprimentos de onda, normalmente contendo o comprimento de onda em que a fibra operará. É' efetuado em laboratório devido à complexidade e precisão e fornece dados sobre a contaminação que pode ter ocorrido na fabricação da pré-forma e puxamento, principalmente o OH^-.

O teste consiste em se medir a potência de luz após percorrer toda a fibra nos vários comprimentos de onda em que se deseja medir a atenuação, esta é a primeira medida, ou ainda, a potência de saída. Após isso, corta-se a fibra a 2 ou 3 metros do início, sem alterar as condições de lançamento, e mede-se a potência de luz nesse ponto, que pode ser considerado como a potência de entrada, uma vez que 2 ou 3 metros causam atenuação desprezível; esta é a segunda medida. De posse das duas medidas, calcula-se a atenuação por

$$A = 10 \cdot \log \frac{P_s}{P_e} \text{ [dB].}$$

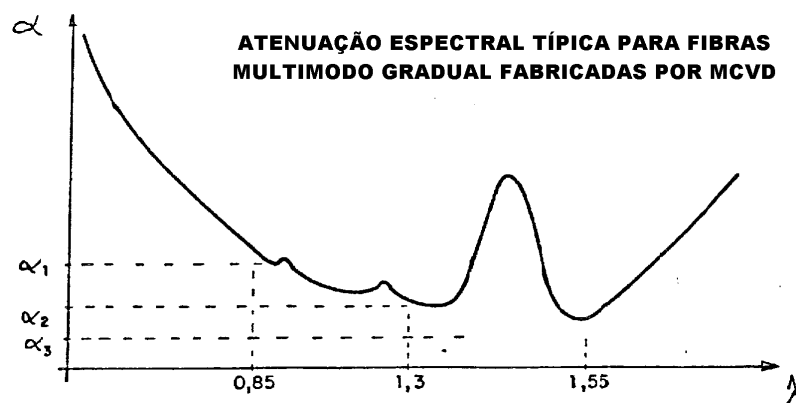



Figura 4-10.1: Atenuação espectral típica para fibras multimodo gradual fabricadas por mcvd

4.10.2. Teste de Atenuação de Inserção

Este teste é mais apropriado para situações de campo e mede a atenuação da fibra óptica apenas num comprimento de onda. Normalmente mede-se no comprimento de onda em que o sistema opera. O teste utiliza dois instrumentos portáteis: o medidor de potência e a fonte de luz.

O teste divide-se em duas etapas, na primeira é efetuada uma calibração dos dois instrumentos, para conhecermos a potência de luz que será lançada na fibra óptica. Na segunda é efetuada a medida de potência após a luz percorrer toda a fibra óptica. A diferença entre as duas será o valor de atenuação.

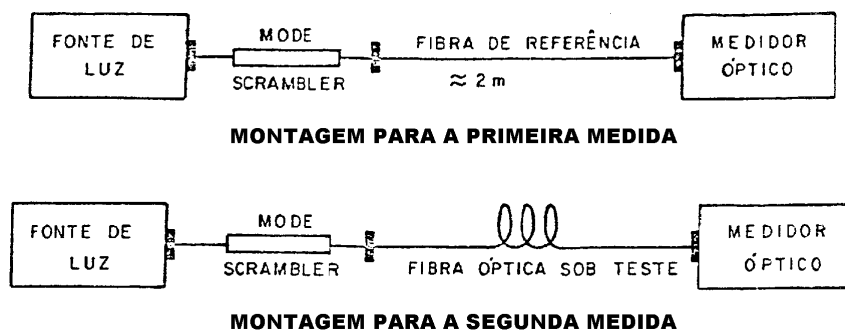



Figura 4-10.2: Teste para atenuação de inserção

4.10.3. Teste de Largura de Banda

Este teste determina a máxima velocidade de transmissão de sinais que uma fibra óptica pode ter, ou seja, mede a capacidade de resposta da fibra óptica. O teste é realizado com o objetivo de saber se a fibra óptica tem condições de operar com a taxa de transmissão especificada para o sistema.

4.10.4. Teste de Abertura Numérica

A abertura numérica é um número que define a capacidade de captação luminosa da fibra óptica e é definida por: $AN = \sqrt{n_1^2 - n_2^2}$, onde:

- n_1 é o índice de refração do núcleo
- n_2 é o índice de refração da casca

ou ainda: $AN = \text{sen}\,\frac{\alpha}{2}$, onde:

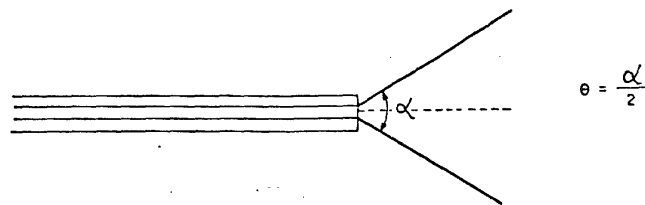



Figura 4-10.3: Teste de abertura numérica

Esta grandeza é intrínseca à própria fibra e é definida na fabricação, onde tem maior importância.

Como a abertura numérica é equivalente à distribuição de luz do campo distante, o teste mede a intensidade de luz desse campo.

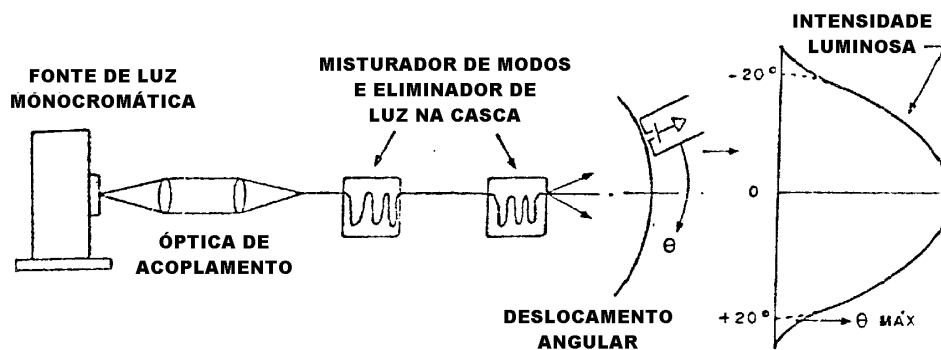



Figura 4-10.4: Medida da distribuição de luz no campo distante

As medidas são obtidas através de um detector que percorre um deslocamento angular ou pela projeção do feixe de luz num anteparo graduado. Desta maneira se obtem o ângulo de abertura do feixe luminoso.

4.10.5. Teste de Perfil de Índice de Refração

Este teste tem maior importância na fase de fabricação de fibras ópticas.

Não existem limites para o perfil de índice, uma vez que qualquer imperfeição no mesmo implica numa diminuição da banda passante da fibra óptica, esta sim com limites específicos. O valor do índice de refração num determinado ponto é proporcional à distribuição de luz do campo próximo.

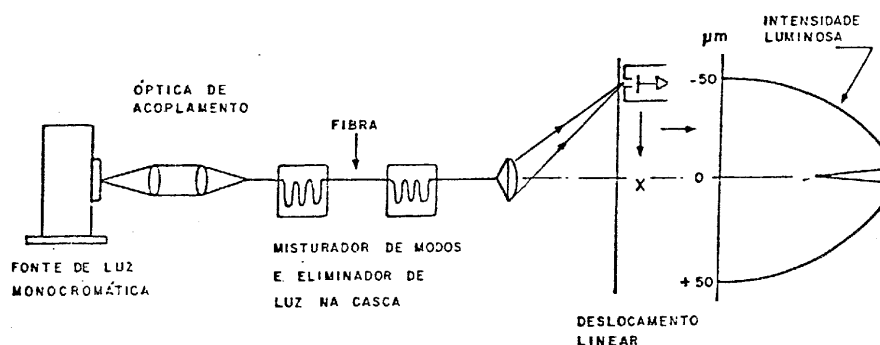



Figura 4-10.5: Medida da distribuição de luz no campo próximo

4.11. FONTES ÓPTICAS

4.11.1. Tipos de Fontes Ópticas

Para sistemas ópticos, encontramos dois tipos de fontes ópticas freqüentemente utilizadas: LED e LASER. As principais características diferenciadoras dessas duas fontes são:

- Potência luminosa:os lasers oferecem maior potência óptica se comparados com os leds.
  - LED : (-7 a -14dBm)
  - LASER : (1dBm)
- Largura espectral: os lasers tem largura espectral menor que os leds, o que proporciona menor dispersão material.

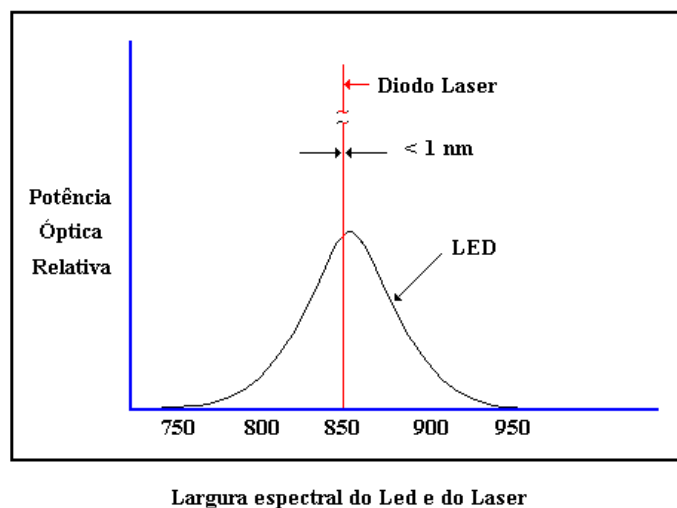



Figura 4-11.1: Largura espectral do LED e do laser

- Tipos e velocidades de modulação: os lasers têm velocidade maior que os leds, mas necessitam de circuitos complexos para manter uma boa linearidade.
- Acoplamento com a fibra óptica: o feixe de luz emitido pelo laser é mais concentrado que o emitido pelo led, permitindo uma eficiência de acoplamento maior.

- Variações com temperatura: os lasers são mais sensíveis que os leds à temperatura.
- Vida útil e degradação: os leds têm vida útil maior que os lasers (aproximadamente 10 vezes mais), além de ter degradação bem definida.
- Custos: os lasers são mais caros que os leds, pois a dificuldade de fabricação é maior.
- Ruídos: os lasers apresentam menos ruídos que os leds. Ambos podem ser fabricados do mesmo material, de acordo com o comprimento onda desejado:
    - AlGaAs (arseneto de alumínio e gálio) para 850 nm.
    - InGaAsP (arseneto fosfeto de índio e gálio) para 1300 e 1550 nm.

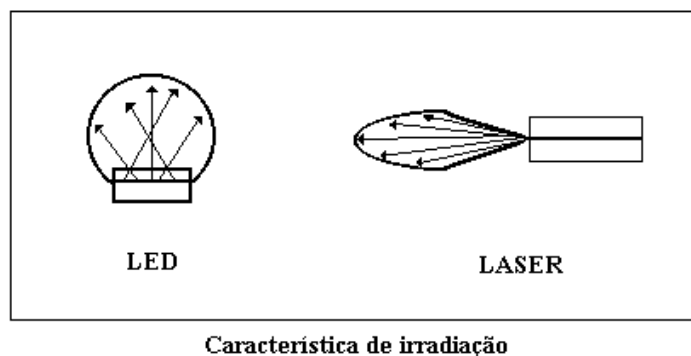



Figura 4-11.2: Características da irradiação

4.11.2. LASER

Para entendermos o funcionamento de um laser, tomemos um laser a gás (HeNe) de maneira didática, onde os números usados são ilusórios para melhor visualização dos fenômenos.

Um átomo é composto de um núcleo e de elétrons que permanecem girando em torno do mesmo em órbitas bem definidas. Quanto mais afastado do núcleo gira o elétron, menor a sua energia. Quando um elétron ganha energia ele muda de sua órbita para uma órbita mais interna, sendo este um estado não natural para o átomo, mas sim forçado. Como esse estado não é natural, por qualquer distúrbio o átomo tende a voltar a seu estado natural, liberando a energia recebida em forma de ondas eletromagnéticas de comprimento de onda definido em função das órbitas do átomo.

As características típicas de um laser são:

- luz coerente
- altas potências
- monocromaticidade
- diagrama de irradiação concentrado
- vida útil baixa (10000 horas)
- sensível a variações de temperatura
- alto custo
- próprio para sinais digitais
- altas velocidades, ou seja, grande banda de passagem (1 Ghz ou mais)

Existem dois tipos de lasers quanto ao tipo de fabricação:

- Lasers cujo guia de onda (cavidade ressonante) é induzida por corrente, chamados lasers GLD (gainguide laser diode).
- Lasers cujo guia de onda é incorporado pela variação de índice de refração, chamados lasers ILD (index guide laser diode).

4.12. REDES ÓPTICAS

Avanços recentes na tecnologia de comunicações ópticas vêm revolucionando a infra-estrutura de telecomunicações em quase todo o mundo. Capacidades de transmissão cada vez mais elevadas têm possibilitado a implantação de redes de comunicações que viabilizam a integração de diferentes tipos de tráfego, a sofisticação de serviços, a interconexão de redes remotas. A transmissão de informação por meio de fibras ópticas para serviços de faixa larga tem se tornado, segundo critérios técnicos e econômicos, a opção mais atraente. Milhões de quilômetros de cabos de fibras ópticas já foram implantados por empresas de televisão a cabo e companhias telefônicas.

A aceitação mundial de sistemas de comunicações baseados no uso de fibras ópticas deve-se, em parte, à padronização internacional da Hierarquia Síncrona Digital (SDH) para transporte, e ao ATM (Modo Assíncrono de Transferência) para o acesso e a comunicação de sinais portadores de informação digitalizada. A adoção destes padrões possibilita, por exemplo, a compatibilização e interoperabilidade de redes heterogêneas.

4.12.1. Arquitetura para redes ópticas

A fibra óptica possui grande potencial para reduzir custos na oferta de largura de faixa, o que pode viabilizar, por exemplo, a oferta de serviços interativos de faixa larga. Serviços tais com videoconferência e distribuição de programação de TV de acordo com as preferências dos usuários podem ser prontamente disponibilizados com a tecnologia atual.

A combinação de alternativas propiciadas pelas novas tecnologias e a busca pelas melhores soluções tanto em termos de arquitetura quanto com relação às especificações de elementos de rede conduziram a várias abordagens com relação ao *loop* de assinantes das redes de telecomunicações. Uma das possíveis arquiteturas para implementação é a FITL (Fiber-in-the-loop), inclusive para a provisão de serviços baseados em vídeo.

Em alguns países, as empresas de TV a cabo foram pioneiras no transporte de sinais de vídeo através de enlaces ópticos via fibra, e grandes incentivadoras do desenvolvimento de tecnologias capazes de transportar sinais de vídeo analógicos multiplexados em freqüência via fibra.

Dentro dessa abordagem, alimentadores a fibra são usados para transportar grandes quantidades de sinais de vídeo (analógicos e digitais) até "nós de distribuição" localizados próximos a áreas com 200 a 2000 usuários. Essa estratégia aproveita os avanços recentes nas tecnologias de modulação e compressão de vídeos digitalizados, hoje capazes de colocar 10 ou mais sinais de vídeo numa largura de faixa de 6 MHz, antes normalmente ocupada por apenas um sinal de vídeo analógico. A partir de nós de distribuição, os sinais de vídeo são distribuídos pelos cabos coaxiais já implantados.

Como os alimentadores a fibra servem grupos com número relativamente pequeno de usuários (quando comparados ao universo de assinantes), os quais, por sua vez, estão ligados a redes de cabos coaxiais com arquiteturas do tipo árvore, o número de amplificadores analógicos (repetidores) na parte coaxial dessas redes fica significativamente reduzido, e a largura de faixa total correspondente a esta parte da rede fica aumentada. O menor número de repetidores na parte coaxial resulta em melhoria da qualidade de transmissão, além de redução dos custos de manutenção. Com a possibilidade de se chegar até o usuário com centenas de canais de TV, pode-se imaginar o fornecimento de programação quase que individualizada.

Uma outra alternativa combina alimentadores a fibra e pares trançados para o trecho entre hubs localizados nas terminações dos alimentadores e os assinantes. Esse esquema é conhecido por Asymetric Digital Subscriber Line ou ADSL. A tecnologia atual permite que se enviem taxas em torno de 1,5 Mbits/s da rede para o assinante via par trançado com até 6 km de comprimento. Para distâncias menores (1,5 a 2 km), pode-se transmitir cerca de 6 Mbits. Isto pode representar uma solução econômica satisfatória para a demanda por aplicações tipo vídeo discado a curto prazo.

A abordagem conhecida por FTTC, ou Fiber to the curb, (fibra até o meio-fio), por sua vez, apesar de semelhante à abordagem descrita no parágrafo anterior, possui alimentadores a fibra que chegam até áreas com 4 a 8 usuários, e os enlaces entre as terminações dos alimentadores e as instalações dos usuários são mais curtos. Tais enlaces podem ser

implementados com par trançado, cabo coaxial ou fibra, dependendo do número de canais de vídeo a serem oferecidos e dos custos comparativos das 3 alternativas.

Em diversos países, empresas telefônicas e de TV a cabo estão implantando redes que empregam sistemas de fibra no enlace do assinante. No caso das redes telefônicas, esses sistemas são normalmente usados para prover tanto serviços tradicionais quanto novos serviços, ao mesmo tempo em que preparam a infra-estrutura para o transporte de televisão e multimídia no futuro.

Nas redes de entroncamentos de centrais locais de comutação telefônica, cada central serve a um certo número de assinantes (tipicamente entre 10.000 e 50.000). A tendência com relação à implantação de redes de entroncamentos sugere um número menor de centrais com grandes capacidades conectadas por anéis a fibra óptica tolerantes a falhas e usando transmissão segundo os padrões do SDH.

Outra arquitetura de rede de faixa larga que emprega ATM e SDH, é a rede de pacotes de faixa larga. Os nós de acesso têm papel semelhante ao de uma central de comutação convencional, com a diferença de que são capazes de realizar comutação de sinais com faixas largas. Nesse tipo de ambiente, toda a infra-estrutura de transmissão baseia-se em tecnologia fotônica.

4.12.2. Ethernet Óptica

Introdução

Fundamentalmente, a Ethernet Óptica representa a combinação e a extensão de duas tecnologias existentes, Ethernet e Sistemas Ópticos. Esta união aproveita o melhor de ambos e amplia suas potencialidades para criar um novo paradigma que fundamentalmente muda a maneira com que os provedores de serviços e as corporações planejam, controlam, e operam suas redes. É uma tecnologia que combina a onipresença, a flexibilidade e a simplicidade do Ethernet com a confiabilidade e a velocidade dos Sistemas Ópticos. Os atributos resultantes da Ethernet Óptica (simplicidade, velocidade, e confiabilidade) removem o estrangulamento da largura de faixa entre a LAN, a MAN e a WAN.

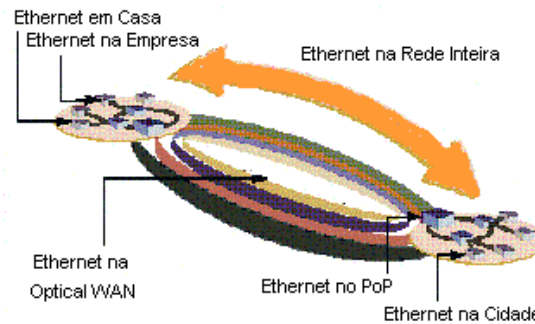



Figura 4-12.1: Ethernet óptica

Ethernet Óptica é o nome dado ao padrão IEEE 802.3ae que suporta a extensão do IEEE 802.3 para taxas de até 10 Gbit/s em redes locais, metropolitanas e de longas distâncias (LANs, MANs, WANs) empregando o método de acesso compartilhado aos meios de transmissão tipo CSMA/CD e o protocolo e o formato de quadro 802.3 do IEEE (Ethernet) para a transmissão de dados.

A Ethernet atual percorreu um longo caminho desde que foi primeiramente implementada nos anos 1970s, onde funcionou originalmente sobre um cabo coaxial espesso e forneceu aos usuários uma conexão compartilhada da largura de faixa de 10 Mbit/s.

A Ethernet logo progrediu e passou a funcionar sobre um par trançado metálico oferecendo conexões dedicadas de 10 Mbit/s usando comutação. Hoje, a Ethernet comutada permite conexões dedicadas de 100 Mbit/s ao computador com troncos de 1 Gbit/s e, dentro de alguns anos, os peritos da indústria predizem 1 Gbit/s ao computador com troncos de 10 Gbit/s.

Em aproximadamente trinta anos de existência, a Ethernet tornou-se onipresente; uma tecnologia plug and play amplamente padronizada, que é usada em mais de noventa por cento das redes locais (LANs) corporativas. A Ethernet alcançou este nível de aceitação porque é simples de usar, barata, e provou seu valor.

Tecnologia de Redes	Ano 2000
Ethernet Total (10 Mbit/s, Fast E, Gigabit E)	91%
Wireless LAN	6%
Token Ring	3%
ATM	<1%

Fonte: IDC, Julho 2001

Tabela 4-12.1: Tecnologias de rede

Do mesmo modo as Tecnologias Ópticas vieram "para ficar" tanto quanto a Ethernet e indiscutivelmente em um período de tempo mais curto.

As velocidades das transmissões ópticas cresceram das dezenas de megabits-por-segundos (Mbit/s) para 40 gigabits-por-segundos (Gbit/s) e os fabricantes de equipamentos ópticos demonstraram recentemente a habilidade de transmitir 6,4 terabits-por-segundos (Tbit/s) sobre um par de fibras ópticas, usando a multiplexação por divisão de comprimento de onda densa (DWDM).

A tremenda capacidade dos Sistemas Ópticos, assim como os avanços em toda a óptica (ou fotônica) e os aperfeiçoamentos das redes ópticas que eliminam a necessidade de regeneração elétrica fixaram os Sistemas Ópticos como a tecnologia de transporte predominante.

A Ethernet Óptica, entretanto, é mais do que apenas a Ethernet aliada aos Sistemas Ópticos. Os participantes dos padrões da indústria assim como os fabricantes têm desenvolvido soluções de Ethernet Óptica específicas, que são mais do que meramente transportar Ethernet sobre Sistemas Ópticos. Em conseqüência, a Ethernet Óptica redefine o desempenho e a economia dos provedores de serviços e das redes corporativas em diversas maneiras.

Atributo Chave de Rede	Redes Atuais	Ethernet Óptica
Complexidade	Maior, múltiplos protocolos da LAN para WAN	Menor, Ethernet da LAN para WAN
Largura de faixa de acesso	Fracionada E1s E3s E4s Dias para provisionar	Até 10 Gbit/s, 2 Mbit/s a qualquer instante Horas para provisionar
Desempenho	Maior latência Maior jitter	Desempenho fim-a-fim da LAN
Pessoal Técnico	Especialistas de rede para cada protocolo	Generalistas de rede

Tabela 4-12.2: Ethernet óptica

Primeiro, a Ethernet Óptica simplifica a rede. Como uma tecnologia sem conexão de camada 2, a Ethernet Óptica remove as complexidades de endereçamento da rede e outros problemas de complexidade da rede, tais como os existentes com as redes baseadas em Frame Relay (FR) e em Asynchronous Transfer Mode (ATM) .

A Ethernet Óptica também remove a necessidade de múltiplas conversões de protocolos que criam grandes dores de cabeça aos operadores da rede com respeito à gerência - porque é Ethernet fim-a-fim. As múltiplas conversões de protocolos também impedem o bom desempenho da rede aumentando a latência e o jitter. Estes atributos são importantes, descrevem o prognóstico e o atraso em uma rede e determinam que serviços e aplicações podem ser oferecidos eficazmente.

A simplicidade da rede Ethernet Óptica também se estende para o provisionamento e a reconfiguração da rede. Os negócios não têm mais que esperar muitos dias por nE1s adicionais ou gastar tempo reconfigurando a rede cada vez que uma mudança seja feita. A largura de faixa da Ethernet Óptica pode ser aumentada ou diminuída sem problemas técnicos e permite mudanças da rede sem reconfigurar cada elemento.

Além da simplicidade, a velocidade é um atributo chave da Ethernet Óptica. Com ela o estrangulamento da largura de faixa é eliminado. A conectividade da Ethernet Óptica permite velocidades de acesso de até 10 Gbit/s (gigabits-por-segundo), ordem de grandeza bem superior que os nE3s atuais. A largura de faixa também está disponível em fatias mais granulares. Os funcionários das equipes de Tecnologia de Informação (TI) não serão mais

forçados a saltar de um E1 a um E3 (2 Mbit/s a 34 Mbit/s) quando tudo o que necessitam realmente é um outro E1 de largura de faixa. Os enlaces de acesso Optical Ethernet podem ser aumentados/diminuídos em incrementos/decrementos de 2 Mbit/s para fornecer a largura de faixa de 2 Mbit/s até 10 Gbit/s ou qualquer valor intermediário.

Além disso, as topologias da Ethernet Óptica permitem uma maior confiabilidade do que as redes de acesso atuais podem fornecer. Por exemplo, os seguintes tipos de soluções de Ethernet Óptica, Ethernet sobre SDH (Synchronous Digital Hierarchy), Ethernet sobre DWDM (Dense Wave Division Multiplex) e Ethernet sobre RPR (Resilient Packet Ring) fornecem a recuperação do tráfego em menos de 50 mili-segundos no evento de uma falha catastrófica, tal como uma interrupção do enlace óptico. Esta confiabilidade elevada garante um tempo útil superior nas redes que fornecem aplicações de missão crítica.

Finalmente, a Ethernet Óptica é significativamente mais barata do que as redes atuais. A economia pode ser vista nos custos operacionais e nos investimentos. O instituto Merrill Lynch, de fato, estima economia de custos de aproximadamente 4:1 para o Gigabit Ethernet contra o ATM (Asynchronous Transfer Mode). Um exemplo simples das economias da infraestrutura vem do fato que as placas de interface Ethernet custam uma fração das placas de interface ATM.

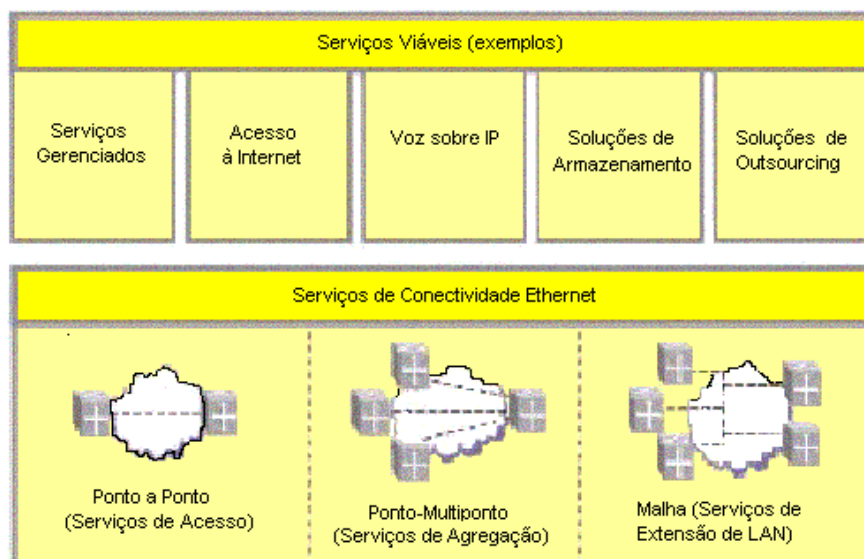



Figura 4-12.2: Exemplos de serviços viáveis e de conectividade Ethernet

As soluções que utilizam a Ethernet Óptica permitem um número extraordinário de novos serviços e aplicações. Estes serviços geralmente são classificados em duas categorias: serviços tipo conectividade Ethernet e serviços viáveis.

Os serviços tipo conectividade Ethernet incluem serviços básicos de Ethernet tais como os de linha privativa e acesso Ethernet, agregação e transporte Ethernet, e extensão de LAN.

Além disso a Ethernet Óptica possibilita uma segunda categoria de serviços e aplicações geralmente chamados serviços "viáveis". Os exemplos destes serviços e aplicações incluem serviços gerenciados (hospedagem de aplicações, desastre/recuperação e soluções de armazenamento) e aplicações específicas da indústria (gerência da cadeia de suprimento, gerência do relacionamento com o Cliente, baseadas em transações e aplicações de comércio eletrônico).

Estes serviços "viáveis" podem incluir qualquer aplicação ou serviço que requer um alto nível de desempenho da rede. Por exemplo, voz sobre IP é uma aplicação idealmente apropriada para o Optical Ethernet pois requer baixos níveis de latência e jitter da rede. Os serviços de armazenamento e desastre/recuperação são outros exemplos de serviços que requerem o desempenho da rede Optical Ethernet. Estes serviços requerem o desempenho em “tempo-real” da rede fim-a-fim, a disponibilidade abundante da largura de faixa, e os mais altos níveis de confiabilidade e segurança da rede.

Adotar a Ethernet Óptica nas suas redes permitirá às operadoras, aos provedores de serviços e às empresas escolherem uma variedade de novos serviços especializados que podem ser rapidamente fornecidos, facilmente provisionados, e mantidos remotamente.

As vantagens provenientes dos novos serviços com a Ethernet Óptica são:

- simplicidade (transparência fim-a-fim);
- velocidade (2Mbit/s até 10Gbit/s);
- confiabilidade (proteções das redes ópticas).

Para as empresas, as possibilidades são emocionantes, e os riscos são minimizados. A Ethernet Óptica fornece a conectividade necessária para permitir aplicações e serviços inovadores que ajudam a maximizar a lucratividade da empresa.

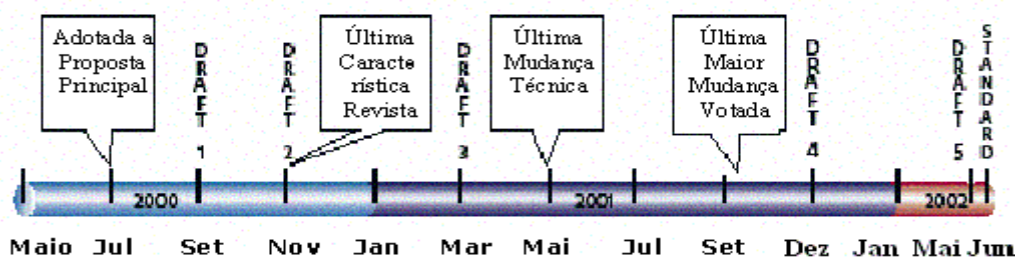
As empresas poderão implementar ou contratar uma nova série de serviços que incluem os seguintes exemplos:

Serviços de conectividade Ethernet: linha alugada Ethernet, Virtual Private Ethernet (VPE), acesso à Internet, acesso ao Internet Data Center (IDC), etc.

Serviços viáveis: recuperação centralizada de desastre/armazenamento, outsourcing de rede, consolidação de aplicação, voz sobre IP, etc.

Padronização

Em junho de 1998, o grupo de trabalho do IEEE 802.3z finalizou e aprovou formalmente o padrão Gigabit Ethernet. Menos de um ano mais tarde, em março de 1999, o grupo de estudo de mais alta velocidade (HSSG) realizou uma "chamada para discussão" para o 10 GE com 140 participantes, representando pelo menos 55 companhias. O grupo HSSG detectou que havia grande necessidade da próxima "velocidade mais alta" da Ethernet baseada em um crescimento rápido da rede e do tráfego da Internet e em uma forte pressão para soluções de 10 Gbit/s, tais como a agregação do GE, os canais da fibra óptica, os roteadores de terabit, e as interfaces de próxima-geração (NGN I/O). Provavelmente a razão mais convincente pela qual o grupo HSSG recomendou ao IEEE a adoção de um padrão de 10 GE foi o seu desejo de evitar a proliferação de Usuários não padronizados, e

Figura 4-12.3: Cronograma de desenvolvimento do padrão 802.3ae

conseqüentemente, prováveis soluções não interoperáveis.

Em janeiro de 2000 o Conselho de Padronização do IEEE aprovou um pedido de autorização de projeto para o 10 GE, e o grupo de trabalho do IEEE 802.3ae começou imediatamente o trabalho. O compromisso com este novo desenvolvimento aumentou consideravelmente, e então mais de 225 participantes, representando pelo menos 100 companhias, foram envolvidos neste esforço técnico.

De fato, um progresso incrível foi alcançado com a proposta inicial do padrão que foi liberado em setembro de 2000 e a versão 2.0 foi liberada em novembro de 2000. Estas primeiras tentativas representaram um marco significativo no processo de desenvolvimento, desde as versões mais pesadamente debatidas. A camada física (PHY) e suas interfaces dependentes dos meios físicos (PMD), foram padronizadas e definidas. O processo de desenvolvimento do IEEE 802.3ae foi realizado com sucesso e alcançou seu objetivo de ser um padrão ratificado em junho de 2002.

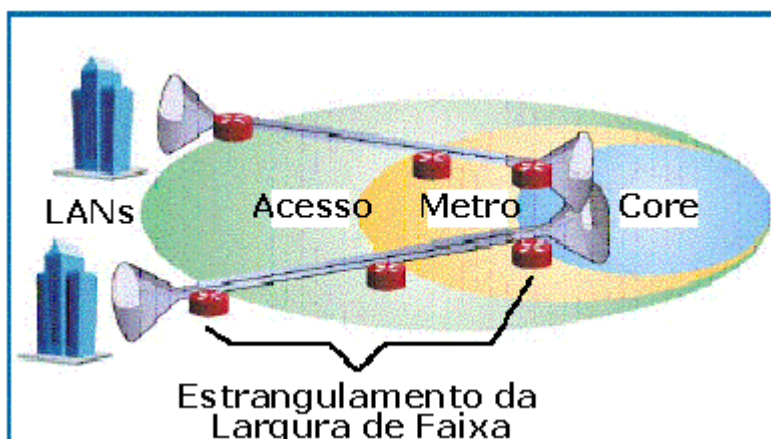



Figura 4-12.4: Estrangulamento da largura da faixa

A Ethernet Óptica fornece hoje o que somente poderia ser imaginado há pouco tempo. Muda fundamentalmente a maneira através da qual as redes estão sendo projetadas, construídas, e operadas criando uma solução nova de interligação que estende os limites do ambiente LAN para abranger a MAN e a WAN.

A Ethernet Óptica fornece um trajeto de transmissão transparente permitindo que os provedores de serviços aumentem seu faturamento e diminuam os custos enquanto continuam a suportar os serviços concedidos.

A Ethernet Óptica permite que as corporações ganhem a vantagem competitiva de suas redes reduzindo seus custos, fornecendo informação mais rápida, aumentando a produtividade dos empregados e melhorando a utilização dos recursos.

A revolução trazida pela Ethernet Óptica será limitada somente por nossas imaginações, fornecendo em uma única solução, uma rede mais rápida, simples, e confiável.

4.13. A FIBRA ÓPTICA E SEUS USOS ATUAIS

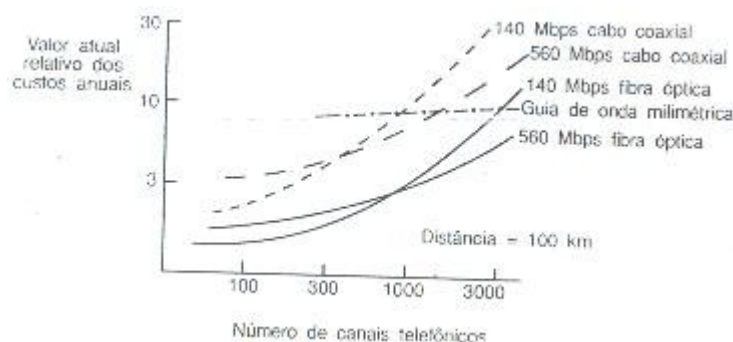
4.13.1. Sistemas de Comunicação

As redes públicas de telecomunicações provêm uma variedade de aplicações para os sistemas de transmissão por fibras ópticas. As aplicações vão desde a pura substituição de cabos metálicos em sistemas de longa distância interligando centrais telefônicas (urbanas e interurbanas) até a implantação de novos serviços de comunicações, por exemplo, para as Redes Digitais de Serviços Integrados (RDSI).

4.13.2. Rede Telefônica

Uma das aplicações pioneiras das fibras ópticas em sistemas de comunicação corresponde aos *sistemas tronco de telefonia*, interligando centrais de tráfego interurbano. Os sistemas tronco exigem sistemas de transmissão (em geral, digitais) de grande capacidade, envolvendo distâncias que vão, tipicamente, desde algumas dezenas até centenas de quilômetros e, eventualmente, em países com dimensões continentais, até milhares de quilômetros. As fibras ópticas, com suas qualidades de grande banda passante e baixa atenuação, atendem perfeitamente a esses requisitos.

A alta capacidade de transmissão e o alcance máximo sem repetidores, permitidos pelos sistemas de transmissão por fibras ópticas, minimizam os custos por circuito telefônico, oferecendo vantagens econômicas significativas. A figura a seguir, por exemplo, compara os custos relativos de sistemas troncos digitais, a 140 Mbps e 560 Mbps, utilizando fibras ópticas e cabos coaxiais. Observe nesta figura que, mesmo com relação aos sistemas em microondas ou ondas milimétricas, as fibras ópticas mostram-se competitivas, pelo menos nos sistemas de maior capacidade.

Figura 4-13.1: Comparação do custo relativo de diferentes meios de transmissão de alta capacidade

Em países ou regiões de intensa urbanização, as distâncias máximas separando centrais ou postos telefônicos são, em geral, inferiores a 100 Km. Nestes casos, os sistemas-tronco de telefonia podem ser implantados, quando necessário, com os repetidores colocados ao longo dos próprios prédios ou instalações telefônicas existentes. Isto evita problemas com a instalação e a alimentação remota dos equipamentos, reduzindo custos e aumentando a confiabilidade e as facilidades de manutenção do sistema.

A interligação de centrais telefônicas urbanas é uma outra aplicação das fibras ópticas em sistemas de comunicações. Embora não envolvam distâncias muito grandes (tipicamente da ordem de 5- 20 Km), estes sistemas usufruem da grande banda passante das fibras ópticas para atender a uma demanda crescente de circuitos telefônicos em uma rede física subterrânea geralmente congestionada. Inúmeros sistemas deste tipo estão instalados deste tipo estão instalados no país e no exterior.

No Japão, desde 1985, está instalado um sistema tronco nacional de telefonia com fibras ópticas, a 400 Mbps, interconectado várias cidades ao longo de um percurso de 3400 Km, com espaçamento entre repetidores de até 30 Km. Com a flexibilidade de expansão permitida pelas fibras ópticas, já está sendo experimentada uma ampliação da capacidade de transmissão do sistema tronco para 1,7 Gbps. Nos EUA, os sistemas-tronco da rede telefônica, instalados até o final de 1987, já consumiram mais de um milhão de quilômetros de fibras ópticas. O espaçamento típico entre repetidores nos sistemas tronco americanos é de 48 Km e a taxa de transmissão é de 417 Mbps, prevendo-se também uma futura expansão da capacidade do sistema para 1,7 Gbps.

4.13.3. Rede Digital de Serviços Interligados (RDSI)

A rede local de assinantes, isto é, a rede física interligando assinantes à central telefônica local, constitui uma importante aplicação potencial de fibras ópticas na rede telefônica. Embora as fibras ópticas não sejam ainda totalmente competitivas com os pares metálicos, a partir da introdução de novos serviços de comunicações (videofone, televisão, dados etc.), através das Redes Digitais de Serviços Integrados (RDSI), o uso de fibras ópticas na rede de assinantes tende a ser imperativo. A Figura abaixo ilustra a evolução no uso de fibras ópticas na rede de assinantes em RDSI.

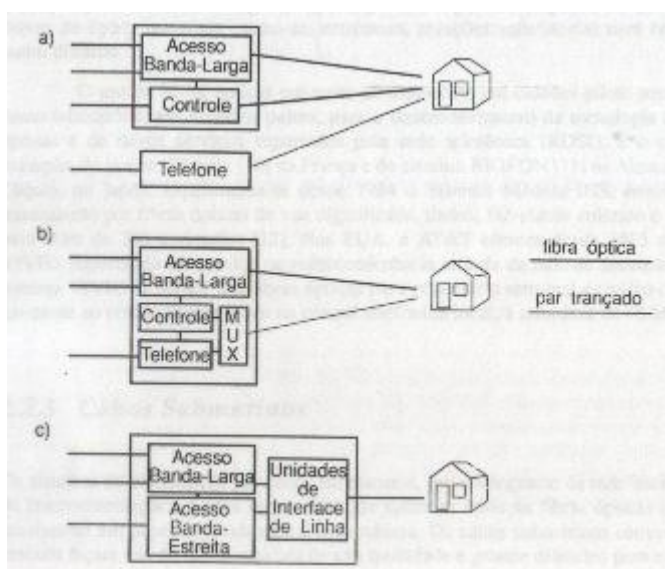



Figura 4-13.2: Evolução da rede de assinantes:
a) fibra óptica para CATV; b) sistema integrado; c) apenas fibra óptica

A grande banda passante oferecida pelas fibras ópticas, além de suportar novos serviços de transmissão permite configurar a rede de assinantes em topologias mais econômicas em termos de cabeação (figura abaixo). Além disso, com as grandes quantidades de fibras ópticas necessárias para a rede de assinantes, o custo dos cabos ópticos deve cair consideravelmente, aproximando-se do custo dos cabos com pares metálicos.

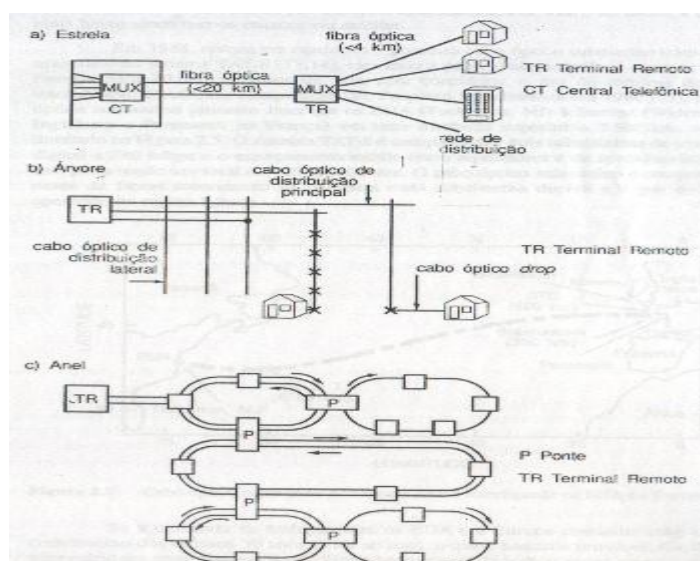



Figura 4-13.3: Arquiteturas para a rede de distribuição dos assinantes com fibras ópticas

Um outro desafio a considerar para a disseminação de fibras ópticas na rede de assinantes, além da redução dos custos atuais, é a interfaces ópticas adequadas aos aparelhos telefônicos. Essas interfaces exigem, por exemplo, a implementação de técnicas para acionamento da campainha e energização do aparelho telefônico via fibra óptica. Com os avanços da tecnologia de componentes optoeletrônicos e dispositivos de óptica integrada espera-se, entretanto, soluções satisfatórias num futuro não muito distante.

O uso de fibras ópticas em redes de assinantes em cidades- piloto tem servido como laboratório, em diversos países, para o desenvolvimento da tecnologia de fibras ópticas e de novos serviços suportados pela rede telefônica (RDSI). É o caso, por exemplo, do projeto Biarritz na França e do sistema BIGFON na Alemanha. Em Tóquio, no Japão, experimenta-se desde 1984 o Sistema Modelo INS, envolvendo a transmissão por fibras ópticas de voz digitalizada, dados, fac-símile colorido e imagem, para mais de 300 assinantes. Nos EUA, a AT&T oferece desde 1985 o sistema VIVID da AT&T usa fibras ópticas para conectar o terminal de vídeo digital do assinante ao comutador de vídeo na central telefônica local, a uma taxa de 45 Mbps.

4.13.4. Cabos Submarinos

Os sistemas de transmissão por cabos submarinos, parte integrante da rede internacional de telecomunicações, é uma outra classe de sistemas onde as fibras ópticas cumprem

atualmente um papel de fundamental importância. As fibras ópticas, por outro lado, considerando-se apenas os sistemas de 3ª geração (1,3μm), permitem atualmente espaçamentos entre repetidores em torno de 60 km.

Com a implantação dos sistemas de transmissão por fibras ópticas de 4ª geração (1,55mm), alcances sem repetidores superiores a 100 Km serão perfeitamente realizáveis. Além disso, as fibras ópticas oferecem facilidades operacionais (dimensões e peso menores) e uma maior capacidade de transmissão, contribuindo significativamente para atender à crescente demanda por circuitos internacionais de voz e de dados, a um custo mais baixo ainda que os enlaces via satélite.

Em 1988, entrou em operação o primeiro cabo óptico submarino transatlântico associado ao sistema TAT-8, elevando a capacidade de tráfego entre os EUA e a Europa para 20.000 circuitos de voz, sem considerar o uso de técnicas digitais de interpolação (TASI) ou compressão. Proposto formalmente em 1980, este cabo óptico submarino pioneiro interliga os EUA (Tuckerton, NJ) à Europa (Widemouth na Inglaterra e Penmarch na França) em uma distância superior a 7500 Km, conforme ilustrado na Figura abaixo. O sistema TAT-8 é composto por dois subsistemas de transmissão digital a 280 Mbps e o espaçamento médio entre repetidores é de aproximadamente 60 Km, perfazendo um total de 125 repetidores. O cabo óptico submarino é composto por 3 pares de fibras monomodo (1 par para cada subsistema duplex e 1 par de reserva) operando na região 1,3 mm.

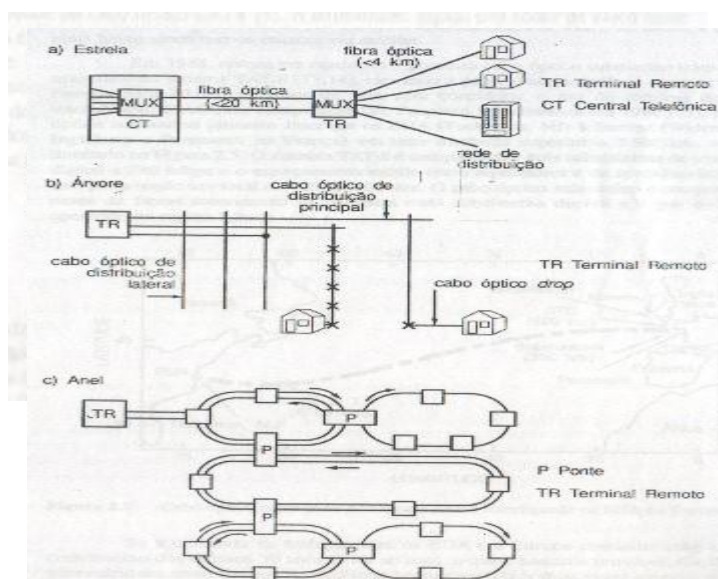



Figura 4-13.4: Cabo óptico submarino TAT-8 interligando os EUA e a Europa

Se a demanda de tráfego entre os EUA e a Europa continuar com a taxa de crescimento dos últimos 30 anos (25% ao ano), o que é bastante provável, em 1992 será necessário um novo sistema transatlântico com capacidade duas vezes superior ao TAT-8. Para enfrentar esta perspectiva, já foi concebido e está desenvolvido o sistema TAT-9, operando em 1,55mm, com maior capacidade de transmissão e espaçamento entre repetidores. O sistema TAT-9 será composto por dois subsistemas a 560 Mbps, interligando, através de unidade de derivação e multiplexação, Manahawkim no EUA e Pennant Point no Canadá a três localidades na Europa (Goonhilly na Inglaterra, Saint Hilaire de riez na França e Conil na Espanha). No total serão 9.000 Kmde cabo óptico submarino com um espaçamento médio entre repetidores da ordem 110 a 120 Km.

No Japão existem atualmente vários sistemas de cabos submarinos com fibras ópticas interligando ilhas do arquipélago, desde sistemas sem repetidores operando nas diferentes hierarquias dos sistemas PCM (32, 6,3 e 1,5 Mbps com fibra índice gradual; 100 e 400 Mbps com fibra monomodo) até um cabo submarino tronco doméstico com repetidores. Os sistemas sem repetidores têm alcances variando de 33 a 48 Km, segundo a taxa de transmissão, e operam a uma profundidade de até 1500 metros. O cabo óptico submarino que compõe o sistema tronco doméstico opera comercialmente desde 1986, a 400 Mbps, com repetidores espaçados de 40 Km, perfazendo um total de 1000 Km a uma profundidade de até 8000 metros.

Na Inglaterra, desde 1987, opera um sistema com cabo óptico submarino, interconectado Dartmouth à ilha de Guernsey no Canal da Mancha, numa distância de 135Km sem repetidores.

Na França, um cabo óptico submarino interliga Marselha no continente a Ajaccio na Córsegaa, numa distância de 330 Km com 9 repetidores.

4.13.5. Televisão a Cabo (CATV)

A transmissão de sinais de vídeo através de fibras ópticas é uma outra classe de aplicações bastante difundida. As fibras ópticas oferecem aos sistemas de CATV, além de uma maior capacidade de transmissão, possibilidades de alcance sem repetidores (amplificadores) superior aos cabos coaxiais banda-larga.

Nos sistemas CATV com cabos coaxiais banda-larga, o espaçamento entre repetidores é da ordem de 1 Km e o número de repetidores é em geral limitado a 10 em função do ruído e distorção, enquanto que com fibras ópticas o alcance sem repetidores pode ser superior a 30 Km. Além de um melhor desempenho, a tecnologia atual de transmissão por fibras ópticas é competitiva economicamente e apresenta uma confiabilidade substancialmente melhor que os sistemas CATV convencionais com cabos coaxiais banda-larga.

Embora a transmissão de imagem digital permita um desempenho superior, os custos dos equipamentos envolvidos com a digitalização têm restringido o uso de fibras ópticas em sistemas CATV com transmissão se sinais de vídeo, principalmente na forma analógica.

Um dos primeiro sistemas comerciais de CATV com fibras ópticas foi instalado em 1976, em Hasting, Inglaterra. Este sistema pioneiro tinha uma extensão de 1,4 Km, distribuindo sinais de vídeo para 34.000 assinantes. Um outro exemplo de sistema pioneiro de transmissão de vídeo por fibras ópticas, neste caso, de transmissão de vídeo digital, é dado pelo sistema instalado na cidade de London (Ontário), Canadá, interligando um estúdio central de distribuição ao conversor de freqüências (head end) na extremidade do cabo tronco CATV. A transmissão digital dos sinais de vídeo neste sistema é feita 322 Mbps, em um cabo óptico com 8 fibras, transportando 12 canais de vídeo e 12 canais FM estéreos numa distância de 7,8 Km. No Japão, um sistema experimental de CATV por fibras ópticas opera a 900 Mbps com 8 canais de vídeo e 16 canais de áudio num tronco de até 20 Km. Grandes avanços neste campo são esperados são esperados com a introdução de multiplexação por divisão em freqüência através dos sistemas ópticos coerentes.

4.13.6. Sistemas de Energia e Transporte

A difusão das fibras ópticas nas redes públicas de telecomunicações tem estimulado a aplicação desse meio de transmissão em sistemas de utilidade pública que provêm suas próprias facilidades de comunicações, tais como os sistemas de geração e distribuição de energia elétrica e os sistemas de transporte ferroviário. As facilidades de comunicações incluem, além de serviços de comunicação telefônica, serviços de telemetria, supervisão e controle ao longo do sistema. As distâncias envolvidas podem ser de alguns quilômetros ao longo de linhas de transmissão ou linhas férreas.

4.13.7. Redes Locais de Computadores

As comunicações entre computadores são suportadas por sistemas de comunicação de dados que costumam ser classificados, segundo as distâncias envolvidas, em redes de computadores de longa distância ou redes locais de computadores.

As *redes de computadores a longa distância* utilizam-se dos meios de transmissão comuns à rede telefônica. Embora geralmente usem técnicas distintas (comutação de pacotes, modem etc.) essas redes a longa distância são implantadas ou integradas nos mesmos suportes físicos de transmissão da rede telefônica. Assim sendo, o uso de fibras ópticas em sistemas de comunicação de dados a longa distância acompanha a evolução da aplicação de fibras ópticas na rede telefônica (cabos troncos, cabos submarinos, RDSI etc.).

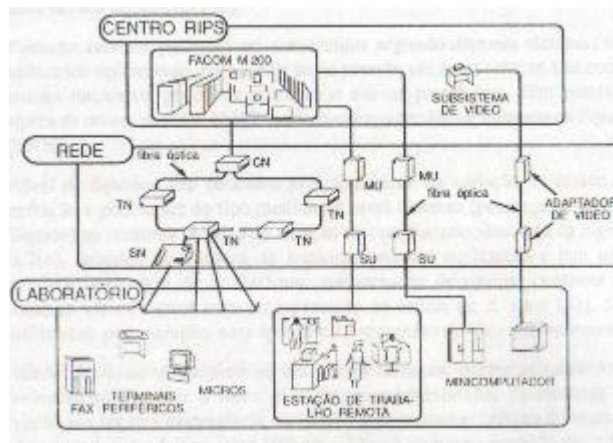
As redes locais de computadores, utilizadas para interconectar recursos computacionais diversos (computadores, periféricos, banco de dados etc.) numa área geograficamente limitada (prédio, usina, fábrica, campus etc.), caracterizam-se pela especificidade e variedade de alternativas tecnológicas quanto ao sistema de transmissão voltadas principalmente para aplicações em automação em escritórios e em automação industrial, como requisitos exigentes em termos de confiabilidade, capacidade de uma excelente alternativa de meio de transmissão.

Sistemas centralizados, envolvendo um computador de grande porte (mainframe) e vários terminais remotos, embora não sendo propriamente uma rede de computadores, constituem sistemas de comunicação de dados em nível local onde a substituição de cabos metálicos por fibras ópticas pode ser vantajosa. Utilizadas em barramentos internos serializados ou na interligação dos terminais remotos, as fibras ópticas permitem, desde um melhor desempenho em termos de alcance e banda passante, até maiores facilidades (menor volume e peso) na instalação dos cabos.

Existem vários exemplos do uso de fibras ópticas em redes locais de computadores. De um modo geral, as iniciativas buscam usufruir uma ou mais qualidades das fibras ópticas a fim de atenderem a situações diversas, tais como, por exemplo:

- necessidade de maior alcance da rede ou de segmentos dentro da rede;
- demanda de novos serviços de comunicação local exigindo grandes bandas passantes;
- necessidade de maior confiabilidade do sistema em ambientes hostis, etc.

Em razão dos custos associados aos nós de comunicação serem ainda relativamente altos, o uso da tecnologia de fibras ópticas em redes locais de computadores tem se limitado principalmente aos grandes sistemas. É o caso, por exemplo, do sistema RIPS (Research Onformation Processing System) do Centro de Pesquisa de Tsukuba no Japão e da rede com integração de voz que a companhia Delta Air Lines opera no aeroporto internacional de Atlanta nos EUA. O sistema RIPS, cuja configuração geral é mostrada na Figura abaixo, integra, através de fibras ópticas, serviços de transmissão de voz, dados e imagem para atender às atividades de P&D de mais de 3.000 pessoas. Uma outra classe de aplicação, justificando economicamente o uso de fibras ópticas em redes locais de computadores, tem sido em fábricas ou plantas onde os processos têm requisitos de confiabilidade imperativos (usinas nucleares, elétricas etc) ou exigem grandes capacidades de transmissão, como os sistemas de manufatura integrada (CAD, CAM, etc.).

Figura 4-13.5: Configuração geral do sistema RIPS

4.14. A FIBRA ÓPTICA E A EMBRATEL

4.14.1. Introdução

Para dar maior flexibilidade ao atendimento de clientes e introduzir novas funções nas suas redes, a Embratel estará implantando em 2004, após testes de laboratório no Centro de Referência Tecnológica (CRT) e de campo, uma série de novas tecnologias, que seguem a tendência mundial de retomada de novas implementações nas redes de telecomunicações.

4.14.2. SDH-NG (SDH de nova geração)

As mesmas interfaces ethernet, fast-ethernet e até mesmo giga-ethernet serão oferecidas como interfaces de rede de acesso por meio da implementação das estruturas de SDH-NG, que poderão servir também para o tráfego legado (Frame Relay, ATM, TDM etc.). O SDH de nova geração terá suporte dos anéis de fibra óptica de acesso. Da mesma forma, o objetivo básico é de prover maior flexibilidade e menores custos aos clientes.

As redes metro-ethernet sobre fibra óptica e as redes SDH de nova geração serão complementares na prestação de serviços baseados em interface ethernet para os usuários finais.

4.14.3. Metro-Ethernet

As interfaces ethernet e fast-ethernet, bem conhecidas dentro dos ambientes dos usuários, e até mesmo interfaces giga-ethernet serão oferecidas nas novas redes de acesso, diretamente suportadas na vasta rede de fibras ópticas da Embratel. O objetivo é atender ao crescente tráfego IP das corporações, possibilitando maior flexibilidade e menores custos para os clientes. Facilidades como variação de banda por demanda e autoprovisionamento são algumas das funções possíveis nessas novas redes.

Planejamento do Sistema Telefônico

5.1. PROCESSO DE PLANEJAMENTO

O processo de planejamento é altamente influenciado por informações externas, tanto em relação à demanda por serviços, como de natureza econômica. Este processo pode ser dividido em três fases principais:

- Projeto de estrutura topológica para a rede: Tem como base a localização dos componentes, sua interligação e restrições de conectividade. Nessa fase são usados métodos de otimização topológica e teoria de grafos. O produto final dessa fase é uma matriz de conectividade, incluindo as locações ideais para as centrais e concentradores.
- Síntese de rede: Utiliza as informações adquiridas na fase anterior para calcular a dimensão ótima para os componentes, sujeita a restrições de grau de serviço e medidas de desempenho (ex.: atraso e probabilidade de perda). Os cálculos são baseados em um modelo de custo linear, já a otimização é feita não linearmente devido às características de restrição. Os problemas analisados nessa fase são roteamento de tráfego e dimensionamento. O produto final dessa fase é o plano de rotas e o conjunto de enlaces lógicos entre os nós de rede.
- Especificação da tecnologia de transmissão: Fase onde se concretiza a implementação da rede ou fase de roteamento, onde os componentes existentes no mercado são utilizados como cabos, sistemas de comutação e transmissão específicos. Os problemas encontrados nessa fase se relacionam, na maioria das vezes, com multicommodities, funções de custo modulares e não lineares associados às restrições de confiabilidade.

O grande objetivo de uma rede de telecomunicações é servir aos assinantes de forma satisfatória por um preço compatível. Novamente um equilíbrio entre desempenho e custo está em jogo. O planejamento de uma central telefônica é afetado pelos seguintes aspectos:

Demanda de Assinantes	Representa a simples vontade de uma pessoa conversar com outra via serviço telefônico.
Facilidades para os Assinantes	Representa os serviços e facilidades oferecidas aos assinantes, como serviço de busca, desvio de chamadas entre outros.
Normas	São regulamentações da ITU-T e praticas da

	Telebrás a serem seguidas.
Estrutura	Deve-se garantir que não existam defeitos no sistema ou que na pior das hipóteses eles sejam bem raros.
Expansão	E importante que em todas as etapas do projeto de uma central telefônica se tenha em mente que ela estará em constante expansão.
Otimização da Rede	Representa a busca pelo melhor arranjo de vias e centrais sempre procurando a melhora do serviço (cobertura e qualidade de transmissão) oferecido aos assinantes.
Tempo de Retenção	E o tempo em que uma ligação ocupa o canal.
Conceito de Trafego	Produto da taxa media de chamadas durante um intervalo de tempo pelo tempo médio de retenção dessas chamadas durante o mesmo intervalo.

Tabela 5-1.1: Termos de planejamento

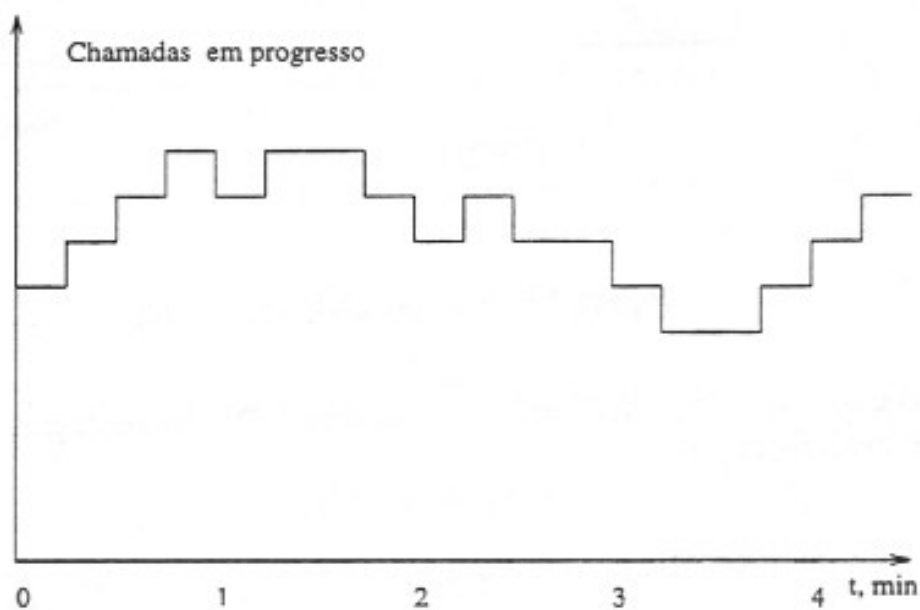



Figura 5-1.1: Conceito de tráfego para curta duração

5.2. TRÁFEGO E CONGESTIONAMENTO

Para projetar os sistemas telefônicos, conforme vimos anteriormente, precisamos levar em consideração uma série de informações, e uma delas é o tráfego telefônico. Precisamos estabelecer o nível do serviço a ser prestado, em função do congestionamento a ser aceito.

5.2.1. A Unidade de Tráfego

A unidade de tráfego erlang (ERL) pode ser definida como o número médio de ocupações que ocorrem simultaneamente em um período de tempo determinado. A. K. Erlang foi um dos pioneiros da Teoria de Tráfego Telefônico. Para órgãos de conversação comuns, o período a ser analisado é geralmente de 1 hora (Rouault, 1976). O tráfego (A), em erlangs, pode ser calculado por:

$$A = \lambda . t_m$$

onde λ é o número de chamadas ou novas ocupações por unidade de tempo, e t_m é a duração média destas chamadas na mesma unidade de tempo.

Existem outros métodos de medição do tráfego. Consideremos que num período T de tempo ocorreram N ocupações, e que a ocupação de um número n de órgãos teve duração t_n. O tráfego é então dado por:

$$A = \frac{1}{T}\sum_{n=1}^{N} t_n$$

Assumindo N canais ou servidores e considerando uma probabilidade B de bloqueio, chegamos ao fator de utilização dos órgãos de saída do sistema telefônico (ρ)

$$\rho = \frac{A.(1-B)}{N},$$

onde A.(1-B) representa o tráfego efetivamente cursado.

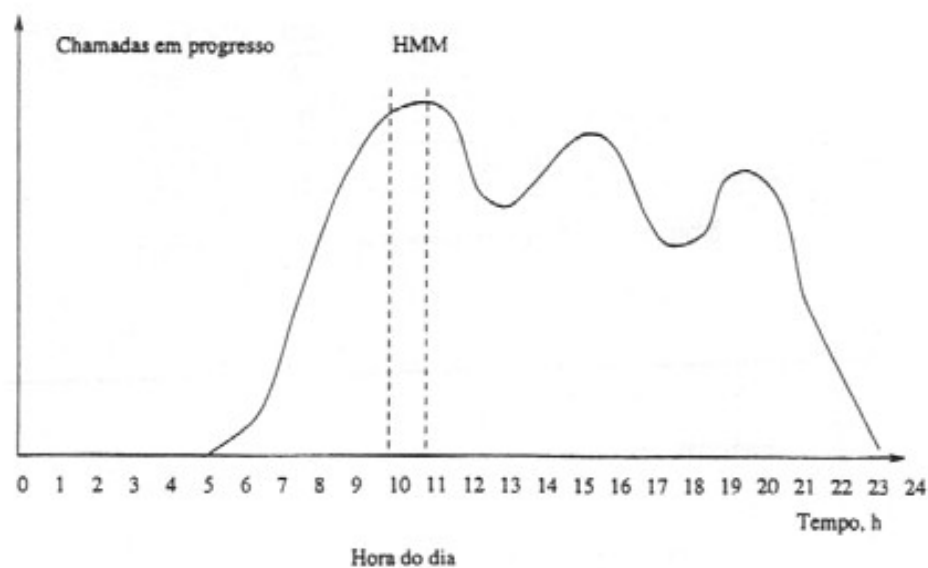



Figura 5-2.1: Tráfego durante o dia

5.2.2. Congestionamento

Por razões obviamente econômicas, o número de órgãos na central telefônica é limitado, ou seja, algumas chamadas são por vezes propositalmente bloqueadas. Isso significa que a chamada pode ser rejeitada (Sistema de Perda) ou tenha que esperar (Sistema de Demora). O conceito de congestionamento pode então ser definido como:

- Congestionamento de Tempo – parte do tempo enquanto os órgãos (ou canais) estão ocupados.
- Congestionamento de Chamadas – parte do número de chamadas que não encontra um canal desocupado.

Todo o processo de medição de tráfego e desempenho nas centrais deve ser feito de forma sistemática, com coleta automatizada de dados, dados esses geralmente obtidos a partir de contadores distribuídos por cada grupo de troncos. Isso permite um acompanhamento em tempo real pela operadora, que pode reconfigurar o sistema de forma a evitar ou minimizar congestionamentos e problemas (Borges, 1985).

Podemos listar algumas medições realizadas pelas centrais que se utilizam de alguns contadores específicos:

- Tráfego por rota de saída e entrada;
- Número de tentativas de chamadas por rota;
- Número de chamadas completadas por rota;

- Tráfego interno;
- Determinação da HMM (hora de maior movimento)

Podemos citar como exemplo os principais indicadores de desempenho do sistema telefônico e objetivo do Sistema Telebrás (Netoet al., 1991) da tabela abaixo:

Indicadores	Sigla	Objetivo
Taxa de solicitações de consertos por 100 telefones	R1	4%
Taxa de solicitações de consertos repetidas em 30 dias	R1.3	15%
Taxa de atendimento de reparo	R2	95%
Taxa de ordens de serviço completadas	S1	95%
Taxa de reclamação de contas	S2	0,6%
Taxa de obtenção de tom de discagem	T1	98%
Taxa de atendimento de serviços especiais	T3	90%
Taxa de chamadas completadas de DDD	T6	58%

Tabela 5-2.1: Indicadores de desempenho do sistema telefônico

5.3. PRINCÍPIOS DO DIMENSIONAMENTO DO SISTEMA TELEFÔNICO

Por Girard, 1990, os princípios básicos para dimensionar o sistema telefônico são:

- As várias partes de um sistema telefônico podem ser ampliadas em etapas
- O tamanho de cada etapa é escolhido de forma a satisfazer as condições determinadas por um período de tempo específico (normalmente 6 meses)
- Para este período, faz-se uma previsão das possíveis variações e crescimento do tráfego.

5.3.1. Fatores que afetam o dimensionamento

Podemos citar como principais fatores que afetam o dimensionamento dos órgãos do sistema telefônico os seguintes:

- Rendimento – dependerá do tráfego efetivamente conduzido, ou do número de conversações estabelecidas.
- Despesa – dependerá do nível do serviço a ser prestado aos assinantes, sob condições de pico de tráfego.
- Volume de Tráfego – a maior parte do volume do tráfego conduzido é escoada quando não há congestionamento. Por isso é importante para o cálculo do rendimento do tráfego.
- Ausência de Congestionamento – o dimensionamento do sistema prevê congestionamento, pois, economicamente, se ele não existe significa que o sistema foi superdimensionado, gerando gastos desnecessários.
- Melhoria no grau de serviço – resulta, geralmente, em um crescimento no tráfego, quando se nota maior facilidade em completar ligações e estabelecer conexões.
- Fatores humanos – sempre causam mais falhas do que o sistema telefônico em si.
- Variações de tráfego – dependem da atividade da comunidade em foco, segundo os seguintes critérios:
    - Picos uma, duas ou três vezes ao dia, num dia de trabalho normal;
    - Varia com as estações do ano;

- Tendência de crescimento, que não é uniforme.

Podemos descrever a variação do tráfego em períodos mais longos (1 ano, por exemplo), através de um histograma. O tempo médio de retenção pode variar tanto diariamente como com as estações do ano, porém menos sensivelmente do que as variações de tráfego.

5.3.2. Estatísticas do Sistema

Podemos utilizar a teoria do tráfego telefônico em cálculos práticos de congestionamento em momentos em que o tráfego encontra-se estacionário, geralmente condições de pico.

No entanto, o tráfego é função do interesse do assinante em efetuar uma ligação, e o perfil deste assinante pode ser levantado através de estatísticas e da elaboração de um modelo sociológico adequado para o usuário típico. A decisão de repetir uma chamada, que influencia e muito no planejamento do sistema, depende de:

- Número de tentativas sem sucesso;
- Incerteza sobre o número discado;
- Possibilidade de ter ocorrido falha no sistema;
- Conhecimento dos hábitos do assinante a que se destina a chamada;
- Grau de urgência da chamada.

Por outro lado, a decisão de desistir depende de:

- Considerar a tentativa inútil;
- Grau de urgência pequeno.

A figura abaixo mostra um exemplo de tipos de tráfego e perdas em um sistema telefônico, onde Ncb representa o número de chamadas básicas, N é o número total de chamadas, incluindo as que foram originadas por repetição, CO são as não completadas por congestionamento, LO as não completadas pela linha estar ocupada, NR as não completadas por conta do assinante B não responder e/ou representando outros motivos para perda (dados da TELERN).

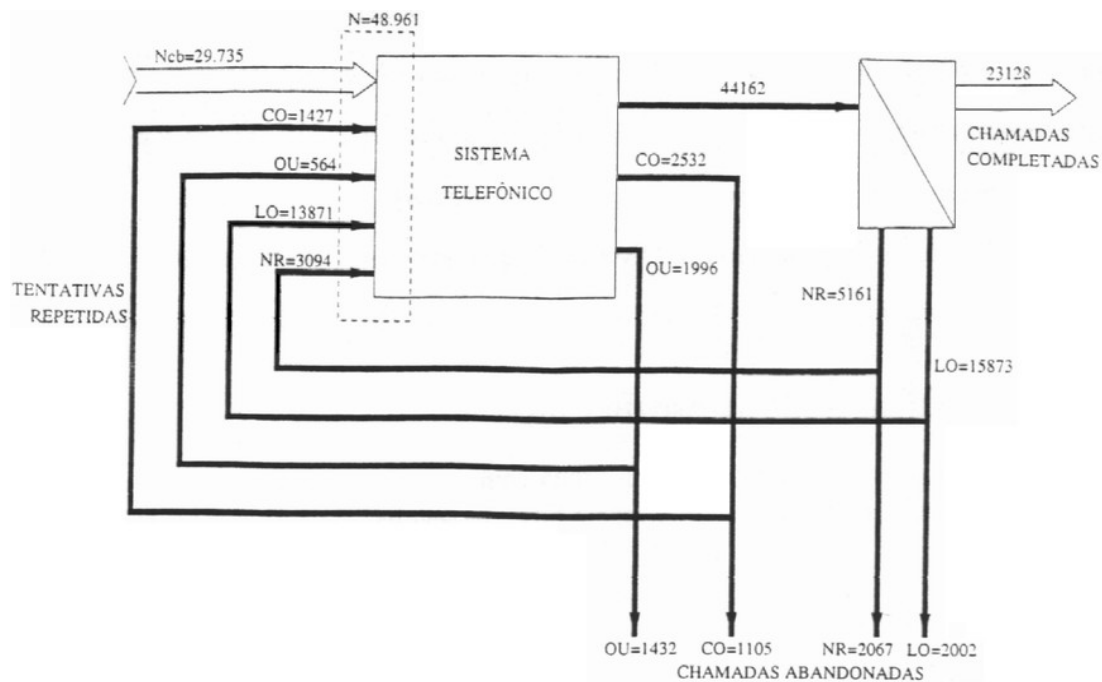



Figura 5-3.1: Exemplo de tráfego em um sistema telefônico

Vemos que de 100 chamadas básicas, apenas 76 são efetivamente completadas. Dessa forma, o planejamento do sistema deve levar esses dados em consideração de forma a incluir procedimentos que minimizem as perdas verificadas.

Podemos ver algumas estatísticas relativas à qualidade do serviço oferecido na Tabela a seguir. Podemos expressar o número médio de tentativas por chamada completada como $N = \frac{1}{1 - q.r^2}$, onde q representa a probabilidade de insucesso numa tentativa e r a probabilidade de repetição de uma tentativa sem sucesso.

Característica	Boa qualidade	Má qualidade
Discagem incompleta ou errada	5-10	13
Congestionamento ou falha	1-5	25
Assinante B ocupado	10-20	20
Assinante B não responde	10-15	10
Não há conversação	30-50	68
Conversação estabelecida	50-70	32
Número de tentativas por chamadas	1,4	4,8

Tabela 5-3.1: Estatísticas do Sistema Telefônico

A próxima tabela apresenta os tempos de ocupação típicos em um sistema telefônico, incluindo o tempo médio de conversação em ligações locais, interurbanas e internacionais e o tempo para uma nova tentativa.

Tipo de Ocupação	Tempo médio (s)
Conversação local	45-120
Conversação interurbana	180-300
Conversação internacional	300-600
Assinante A ouvindo toque de “não responde”	20
Assinante A ouvindo toque de “ocupado”	5
Tempo entre o início da discagem e o ruído de discar	2
Discagem por dígito	1,5
Tempo para uma nova tentativa	30

Tabela 5-3.2: Tempos de ocupação

Como princípios adicionais a serem observados temos:

- A receita operacional da empresa é função direta do tráfego cursado, do número de chamadas completadas;
- O investimento é proporcional ao grau de serviço oferecido ao assinante na hora de maior movimento (HMM);
- Grande parte do tráfego escoa nos intervalos em que não há congestionamento;
- A probabilidade de grandes congestionamentos ocorrerem é baixa;
- Não é economicamente viável projetar o sistema de forma a evitar completamente a ocorrência de congestionamentos;
- Uma melhoria no desempenho do sistema geralmente acarreta uma elevação no tráfego;
- Os assinantes, de um modo geral, não percebem ou não sentem congestionamentos abaixo de 10%;
- O comportamento do assinante pode levar a um desempenho inadequado do sistema.

5.4. CARACTERÍSTICAS DO PROCESSO DE CONEXÃO

Uma conexão é feita a partir de um número de estágios sucessivos. Para calcular e avaliar cada um destes estágios, devemos definir os seguintes fatores (Siemens, 1975):

- Tráfego de entrada para o estágio;
- Agrupamento de estágios;
- Método de busca;
- Sistemas de perda e demora.

Supomos, usualmente, o equilíbrio estatístico, definindo o tráfego como estacionário na média. Podemos então definir:

- Entrada de Bernoulli ou Engset – o número de fontes é finito. A probabilidade de que exatamente uma chamada chegue no intervalo τ e $\tau + \Delta\tau$ é dada por:

$$\omega = \lambda.(M - n).\Delta\tau,$$

onde M é o número de fontes existentes, n é o número de fontes ocupadas, $\lambda\Delta\tau$ é a intensidade de chamadas por fonte quando esta se encontra livre (uma fonte ocupada tem intensidade 0).

- Entrada de Poisson – número infinito de fontes e intensidade de chamadas independente do número de ocupações.

$$\omega = \lim_{M\to\infty,\Delta\tau\to\infty} \lambda.(M-n)\Delta\tau = c_A\Delta\tau,$$

onde $c_A = \lambda M$ é a quantidade média de chamadas oferecidas ao tronco de saída por unidade de tempo e λ representa um fator de proporcionalidade para a intensidade das chamadas das fontes de tráfego.

Para a distribuição dos tempos de retenção adota-se, normalmente, a distribuição exponencial $p_T(t) = \alpha\, e^{-\alpha t}u(t)$, mostrada na Figura 5.6, para dois valores do parâmetro α = 1 e α = 2.

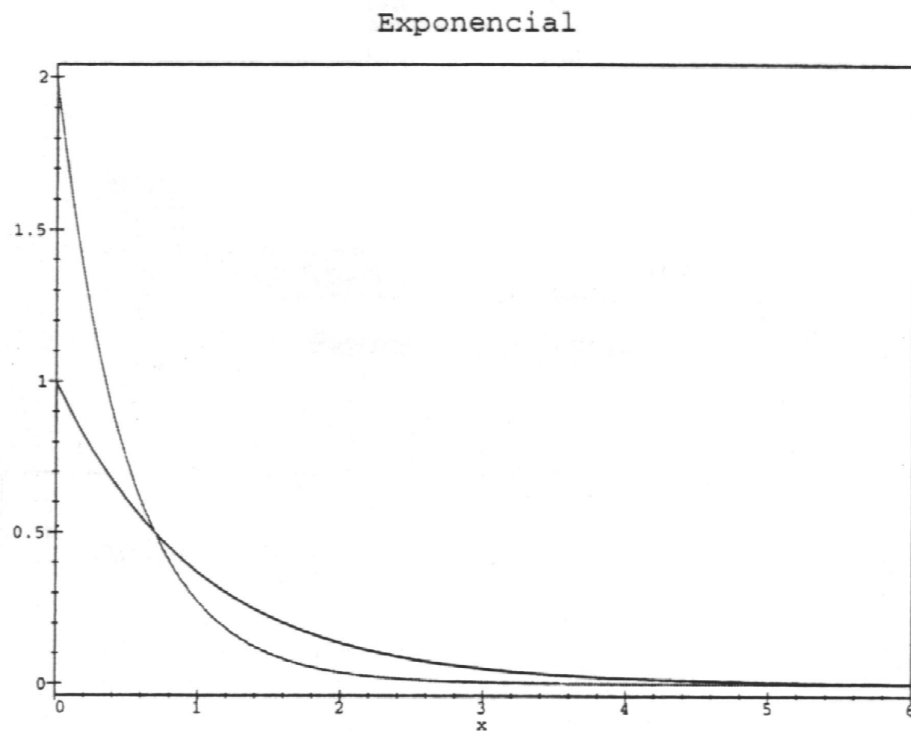



Figura 5-4.1: Distribuição exponencial, para dois valores do parâmetro,
$\alpha = 1$ (curva superior) e $\alpha = 2$ (curva inferior)

Podem ocorrer tempos de retenção constantes ou exponencialmente distribuídos. Mais algumas definições:

- Agrupamentos – grupo de acessibilidade plena: qualquer entrada tem acesso a qualquer saída;
- Graduação – uma entrada tem acesso apenas a um número limitado de saídas. Entradas com acesso às mesmas saídas formam um grupo de entrada. Nenhum grupo de entrada é independente dos outros, pois sempre há saídas comuns a mais grupos;
- Sistema de Enlace – conexão da entrada à saída por meio de um ou mais estágios;
- Sistema de Enlace Graduado – combinação entre Sistema de Enlace e Graduação.

Em relação ao método de busca, podemos considerar:

- Busca seqüencial – sempre a 1ª saída livre é ocupada;
- Busca aleatória – qualquer saída livre pode ser ocupada a qualquer momento.

As características básicas dos sistemas de perda e demora são (Siemens, 1975):

- Sistema de Perdas

As chamadas sem sucesso não provocam acréscimo na intensidade de chamadas bloqueadas;

Chamadas sem sucesso levam a novas tentativas com certa probabilidade.

- Sistema de Demora

Chamadas sem sucesso esperam até serem atendidas (modelo mais simples matematicamente);

Chamadas sem sucesso são abandonadas com uma certa probabilidade;

Chamadas sem sucesso esperam um tempo máximo.

Quanto à disciplina das filas, os métodos de escoamento são:

- Fila Ordenada – A 1ª chamada que chega é servida;
- Fila Aleatória – Cada chamada da fila tem a mesma probabilidade de ser servida, independente do tempo de espera acumulado;
- Fila Prioritária – Cada chamada tem uma prioridade e será atendida de acordo com este critério.

Redes de Computadores

6.1. INTRODUÇÃO A REDES DE COMPUTADORES

6.1.1 Conceito de rede

Uma *rede* é um sistema de pessoas ou objetos conectados entre si. As *redes* estão em todo lugar, até mesmo dentro de nós. Nossos próprios sistemas nervoso e cardiovascular são *redes*. O diagrama abaixo figura mostra vários tipos de *redes*; você pode se lembrar de outros. Observe os agrupamentos:

- Comunicações
- Transporte
- Social
- Biológico
- Serviços públicos
- Outros mais...

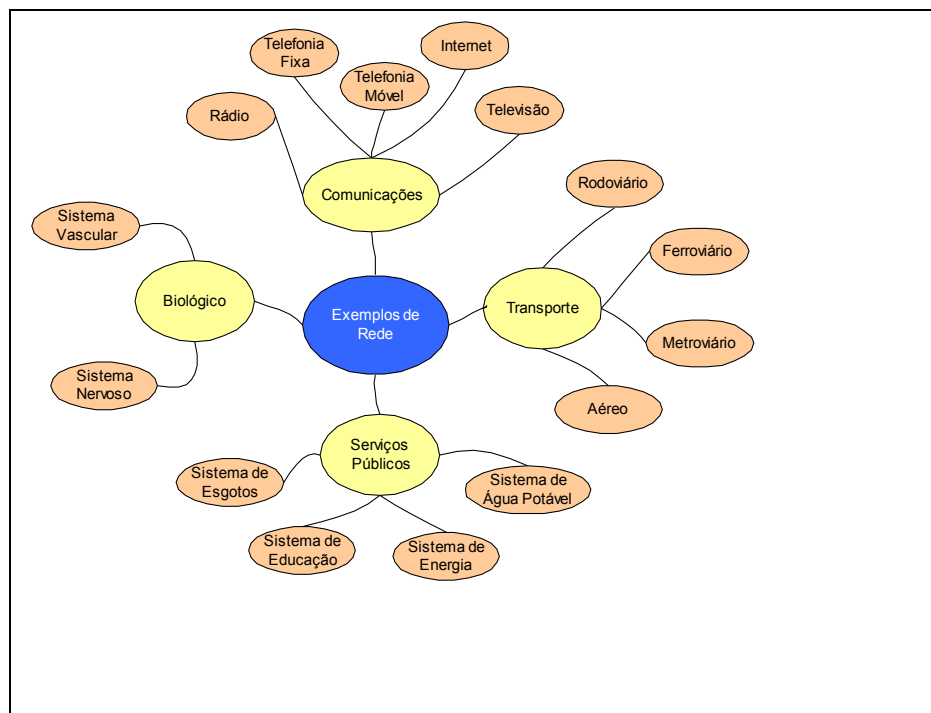



Figura 6-3.1: Exemplos de Rede

6.1.2. O surgimento das redes de computadores

As redes de dados surgiram como resultado de aplicativos de computador que foram criados para empresas. Entretanto, na época em que esses aplicativos foram criados, as empresas possuíam computadores que eram dispositivos dedicados, e cada um deles operava independentemente de outros computadores. Portanto, percebeu-se que essa maneira de administrar empresas não era eficaz nem econômica. Elas precisavam de uma solução que respondesse satisfatoriamente às três questões abaixo:

1. Como evitar a duplicação de equipamentos e recursos
2. Como se comunicar eficazmente
3. Como configurar e gerenciar uma rede

As empresas reconheceram o quanto poderiam economizar e ganhar em produtividade usando a tecnologia de rede. Elas começaram a implantar redes e a expandir as redes existentes quase tão rapidamente quanto surgiam novos produtos e tecnologias de rede. Como resultado, houve uma grande expansão das redes no começo dos anos 80; entretanto, o início do desenvolvimento das redes foi caótico de muitas maneiras.

Em meados dos anos 80, sentiram-se os problemas do crescimento. Muitas das tecnologias de rede que surgiram tinham sido criadas usando-se diferentes implementações de hardware e software. Em conseqüência, muitas das novas tecnologias de rede eram incompatíveis umas com as outras. Tornou-se cada vez mais difícil paras as redes que usavam especificações diferentes se comunicarem entre si.

Uma das primeiras respostas a esses problemas foi a criação de redes locais (LANs). Como elas podiam conectar todas as estações de trabalho, periféricos, terminais e outros dispositivos em um único edifício, as LANs tornaram possível às empresas que usavam a tecnologia da computação, compartilhar, por exemplo, arquivos e impressoras, de modo eficiente.

À medida que o uso do computador nas empresas cresceu, logo se percebeu que até mesmo as LANs não eram suficientes. Em um sistema de LAN, cada departamento ou empresa era uma espécie de ilha eletrônica.

Era necessário um modo de passar informações de maneira rápida e eficiente, não só dentro da empresa, mas também de uma empresa a outra. A solução, então, foi a criação de *redes de áreas metropolitanas (MANs)* e de redes de longa distância (WANs). Como as WANs podiam conectar as redes usuárias dentro de grandes áreas geográficas, elas tornaram possível a comunicação entre empresas a grandes distâncias.

6.1.3. Equipamentos de redes de computadores

Antes de nos aprofundarmos no estudo de redes, faz-se necessária a apresentação dos equipamentos básicos de rede, que possibilitam a conexão entre os elementos da mesma.

As redes locais (LANs) consistem em computadores, placas de rede, meios de rede, dispositivos de controle de tráfego de rede e dispositivos periféricos. As LANs permitem que as empresas que usam a tecnologia da computação compartilhem, de modo eficaz, itens como arquivos e impressoras e usem meios de comunicação como correio eletrônico. Elas reúnem: dados, comunicações, computação e servidores de arquivos.

As LANs são projetadas para executar as seguintes ações:

- Operar dentro de uma área geográfica limitada
- Permitir que muitos usuários acessem meios de grande largura de banda
- Fornecer conectividade ininterrupta aos serviços locais
- Conectar dispositivos fisicamente adjacentes

Hub

A finalidade de um hub é gerar os sinais da rede novamente e os retemporizar. Isso é feito no nível de bit para um grande número de hosts (por exemplo, 4, 8 ou mesmo 24), usando um processo conhecido como concentração. Você vai observar que essa definição é muito similar a dos repetidores, e por essa razão um hub também é conhecido como repetidor multiportas. A diferença é o número de cabos que se conectam ao dispositivo. Dois motivos para se usar os hubs: criar um ponto de conexão central para os meios de cabeamento e aumentar a confiabilidade da rede. A confiabilidade da rede é aumentada permitindo-se que qualquer cabo falhe sem afetar toda a rede. Isso difere da topologia de barramento onde, se houver uma falha no cabo, toda a rede será afetada. Os hubs são considerados dispositivos da

camada 1 do Modelo OSI, como veremos a seguir, porque apenas regeneram o sinal e o transmitem por todas as suas portas (conexões da rede).

Bridge

Uma bridge é um dispositivo da camada 2 do Modelo OSI projetada para conectar dois segmentos da LAN. A finalidade de uma bridge é filtrar o tráfego em uma LAN, para manter local o tráfego local e, ainda assim, permitir a conectividade com outras partes (segmentos) da LAN para o tráfego para elas direcionado. Você pode perguntar-se, então, como a bridge sabe qual tráfego é local e qual não é. A resposta é a mesma que o serviço postal usa quando perguntado como sabe qual correspondência é local. Ele olha o endereço local. Cada dispositivo de rede tem um endereço MAC exclusivo na placa de rede, a bridge mantém registros dos endereços MAC que estão em cada lado da bridge e toma essas decisões com base nesse endereço MAC.

Switch Ethernet

Um switch é um dispositivo da camada 2 do Modelo OSI, assim como a bridge. Na verdade, um switch é chamado de bridge multiporta, assim como um hub é chamado de repetidor multiporta. A diferença entre o hub e o switch é que os switches tomam as decisões com base nos endereços MAC e os hubs não tomam nenhuma decisão. Devido às decisões que os switches tomam, eles tornam uma LAN muito mais eficiente. Eles fazem isso "comutando" os dados apenas pela porta à qual o host apropriado está conectado. Ao contrário, um hub enviará os dados por todas as portas para que todos os hosts tenham que ver e processar (aceitar ou rejeitar) todos os dados.

Os switches, à primeira vista, se parecem com os hubs. Os hubs e os switches têm muitas portas de conexão, uma vez que parte de suas funções é a concentração da conectividade (permitindo que muitos dispositivos sejam conectados a um ponto na rede). A diferença entre um hub e um switch é o que acontece dentro do dispositivo. A finalidade de um switch é concentrar a conectividade, ao mesmo tempo tornando a transmissão de dados mais eficiente. Por ora, pense no switch como algo capaz de combinar a conectividade de um hub com a regulamentação do tráfego de uma bridge em cada porta. Ele comuta os quadros das portas de entrada (interfaces) para as portas de saída, enquanto fornece a cada porta a largura de banda completa.

Roteador

O roteador é um dispositivo que está na camada de rede do modelo OSI, também conhecida como camada 3. Trabalhar na camada 3 permite que o roteador tome decisões com base nos grupos de endereços de rede (classes), ao contrário dos endereços MAC individuais da camada 2. Os roteadores também podem conectar diferentes tecnologias da camada 2, como Ethernet, Token-ring e FDDI. No entanto, devido à sua habilidade de rotear pacotes com base nas informações da camada 3, os roteadores se tornaram o backbone da Internet, executando o protocolo IP.

A finalidade de um roteador é examinar os pacotes de entrada (dados da camada 3), escolher o melhor caminho para eles através da rede e depois comutar os pacotes para a porta de saída apropriada. Os roteadores são os dispositivos de controle de tráfego mais importantes nas grandes redes. Eles permitem que praticamente qualquer tipo de computador se comunique com qualquer outro computador em qualquer parte do mundo.

Equipamentos de WAN

Os equipamentos de WAN são destinados a fazer a comunicação entre redes que ocupam uma área geográfica extensa, e por isso têm características próprias que possibilitam essa comunicação. Na Figura 6-1.3 temos alguns exemplos destes equipamentos.

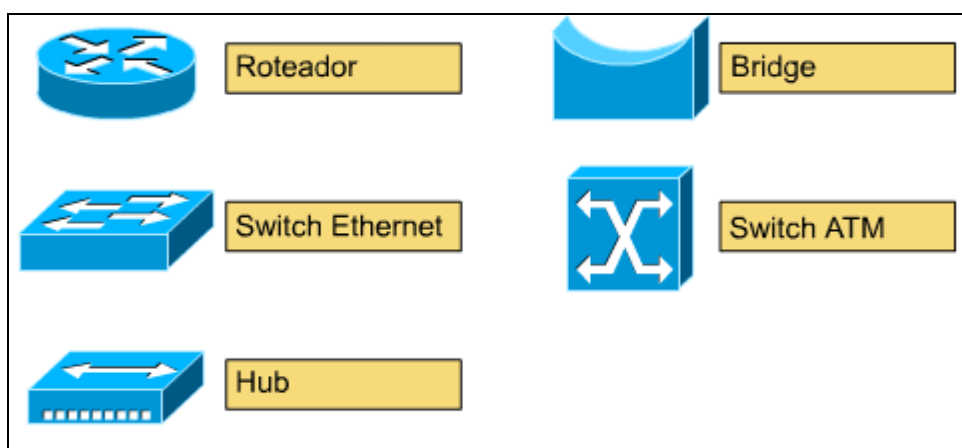



Figura 6-1.2: Elementos de LAN (Padrão CISCO)

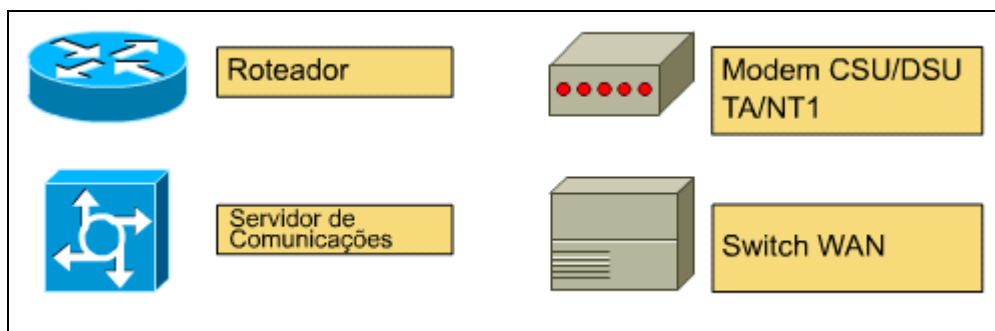



Figura 6-1.3: Elementos de WAN (Padrão CISCO)

6.2. TOPOLOGIAS DE REDE

O termo *topologia* pode ser considerado como "o estudo do local". A topologia é um tema de estudo em matemática, onde os mapas de nós (pontos) e links (linhas) normalmente contêm padrões. Agora vamos examinar as diversas topologias usadas em redes a partir de uma perspectiva matemática. Depois, vamos aprender como uma *topologia física* descreve o plano para cabear os dispositivos físicos. Finalmente, veremos *topologia lógica* para ver como as informações fluem por uma rede para determinar onde as colisões podem ocorrer. Uma rede pode ter um tipo de topologia física e um tipo completamente diferente de topologia lógica. Por exemplo, a Ethernet 10Base-T usa uma topologia física em estrela estendida, mas atua como se usasse uma topologia em barramento lógica. A Token Ring usa uma estrela física e um anel lógico. A FDDI usa um anel físico e lógico.

6.2.1. Topologia de barramento

A *topologia de barramento* tem todos os nós conectados diretamente a um link e não tem outras conexões entre os nós. Cada host é conectado a um fio comum. Nessa topologia, os dispositivos-chave são aqueles que permitem ao host unir-se ou conectar-se ao único meio compartilhado. Uma vantagem dessa topologia é que todos os hosts estão conectados uns aos outros e, portanto, podem comunicar-se diretamente. Uma desvantagem dessa topologia é que um rompimento no cabo desconecta os hosts uns dos outros.

Uma topologia de barramento permite que todos os dispositivos de rede vejam todos os sinais de todos os outros dispositivos. Isso pode ser uma vantagem se você desejar que todas as informações vão para todos os dispositivos. No entanto, isso pode ser uma desvantagem porque os problemas de tráfego e colisões são comuns.

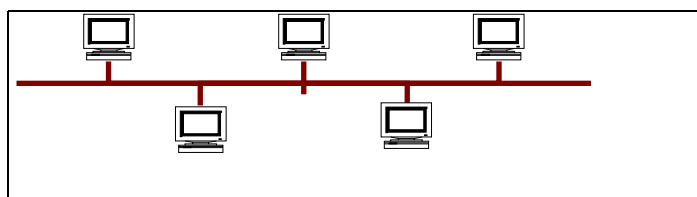



Figura 6-2.1: Topologia em Barramento

6.2.2. Topologia em anel

Uma *topologia em anel* é um único anel fechado que consiste em nós e links, com cada nó conectado a apenas dois nós adjacentes. A topologia mostra todos os dispositivos conectados diretamente uns aos outros, o que é chamado de interligação de equipamentos em cascata.

Para que as informações fluam, cada estação tem de passar as informações à sua estação adjacente.

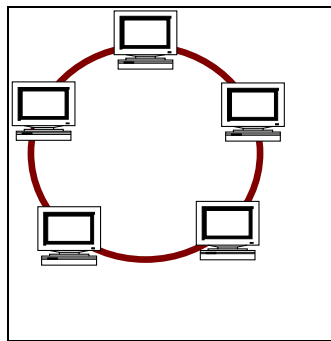



Figura 6-2.2: Topologia em anel

6.2.3. Topologia em anel duplo

A *topologia em anel duplo* consiste em dois anéis concêntricos, cada um conectado apenas ao seu vizinho no anel. Os dois anéis não estão conectados.

Uma topologia em anel duplo é igual a uma topologia em anel, exceto pelo fato de haver um segundo anel redundante que se conecta aos mesmos dispositivos. Em outras palavras, com o objetivo de fornecer confiabilidade e flexibilidade à rede, cada dispositivo de rede é parte de duas topologias em anel independentes.

Uma topologia em anel duplo funciona como dois anéis independentes, dos quais apenas um é usado de cada vez.

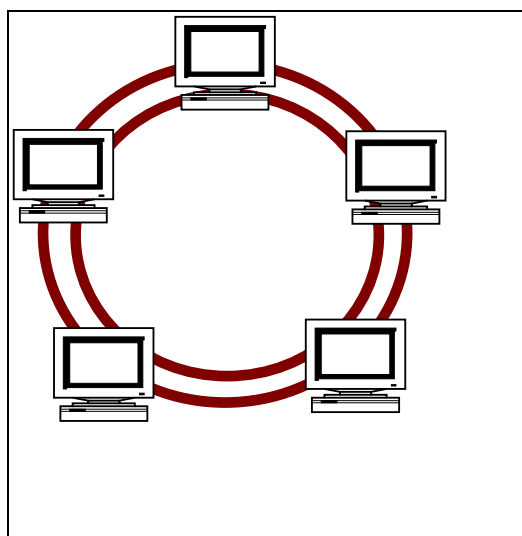



Figura 6-2.3: Topologia em Anel Duplo

6.2.4. Topologia em estrela

A *topologia em estrela* tem um nó central ao qual estão ligados todos os outros nós. Ela não permite outros links.

A principal vantagem deste tipo de estrutura é permitir que os nós se comuniquem uns com os outros mesmo com falha em algum deles. Sua principal desvantagem é que se o nó central falhar, a rede inteira fica desconectada. Dependendo do tipo de dispositivo de rede usado no centro da rede em estrela (Hub/Switch), as colisões podem ser problemáticas.

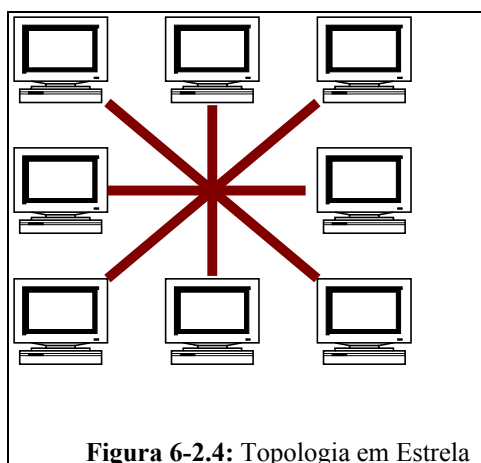



Figura 6-2.4: Topologia em Estrela

O fluxo de todas as informações passa por um único dispositivo. Isso pode ser desejável por razões de segurança ou de restrição de acesso, mas é muito suscetível a qualquer problema no nó central da estrela.

6.2.5. Topologia em estrela estendida

A *topologia em estrela estendida* é igual a uma topologia em estrela, exceto pelo fato de que cada nó vinculado ao nó central é, também, o centro de outra estrela.

Uma topologia em rede estendida tem uma topologia em estrela central, em que cada um dos nós finais da topologia central atua como centro de sua própria topologia em estrela. A vantagem disso é que ela permite que os cabos sejam mais curtos e limita o número de dispositivos que precisam se conectar ao nó central.

Uma topologia em estrela estendida é muito hierárquica, e as informações são encorajadas a permanecerem locais. É assim que o sistema telefônico está estruturado atualmente.

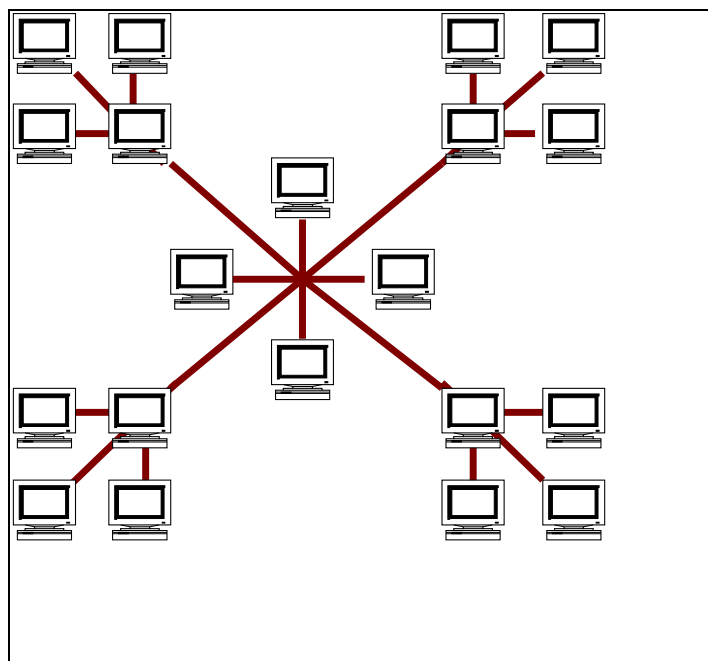



Figura 6-2.5: Topologia em Estrela Estendida

6.2.6. Topologia em árvore

A *topologia em árvore* é similar à topologia em estrela estendida, a principal diferença é que ela não usa um nó central. Em vez disso, ela usa um tronco que se ramifica até outros nós. Há dois tipos de topologias em árvore: a árvore binária (cada nó se divide em dois links)

e a árvore de backbone (um tronco de backbone tem ramos com links pendurados). O tronco é um fio que tem diversas camadas de ramos. O fluxo de informações é hierárquico.

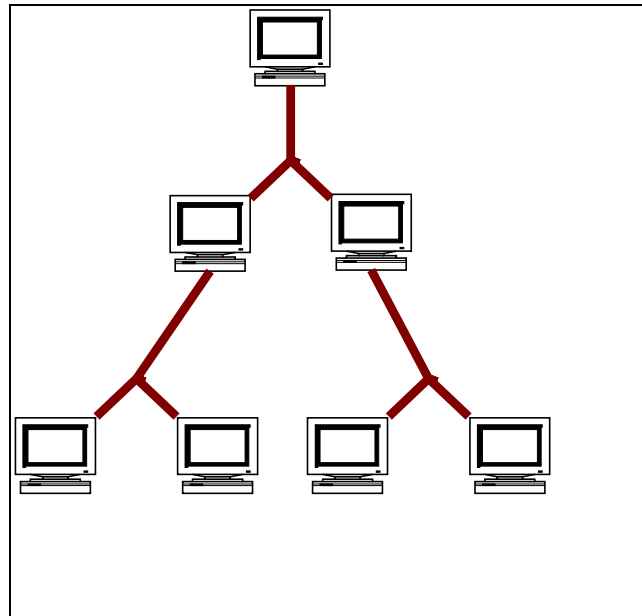



Figura 6-2.6: Topologia em Árvore

6.2.7. Topologia irregular

Na *topologia de rede irregular* não há nenhum padrão óbvio para os links e nós. O cabeamento é inconsistente. Os nós têm números variáveis de fios que partem deles. Essa é a forma como as redes que estão nas etapas iniciais de construção, ou que foram mal planejadas, são freqüentemente cabeadas. Não há nenhum padrão óbvio para os links e nós.

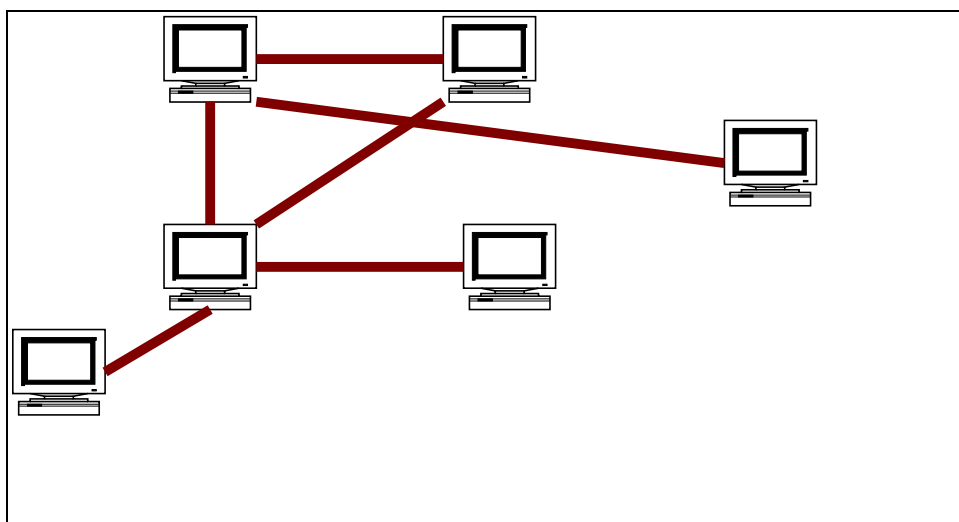



Figura 6-2.7: Topologia Irregular

6.2.8. Topologia em malha

Em uma topologia completa ou *em malha*, cada nó é vinculado diretamente a todos os outros nós. Esse cabeamento tem vantagens e desvantagens muito distintas.

Uma vantagem é que todos os nós estão fisicamente conectados a todos os outros nós (criando uma conexão redundante). Se algum link falhar, as informações poderão fluir através de muitos outros links para atingir seu destino. Outra vantagem dessa topologia é que ela permite que as respostas sejam transmitidas por muitos caminhos de volta através da rede. A principal desvantagem física é que, para um pouco mais que um número pequeno de nós, a quantidade de meios para os links e a quantidade de conexões feitas aos links serão esmagadoras. O comportamento de uma topologia completa, ou malha, depende muito dos dispositivos usados.

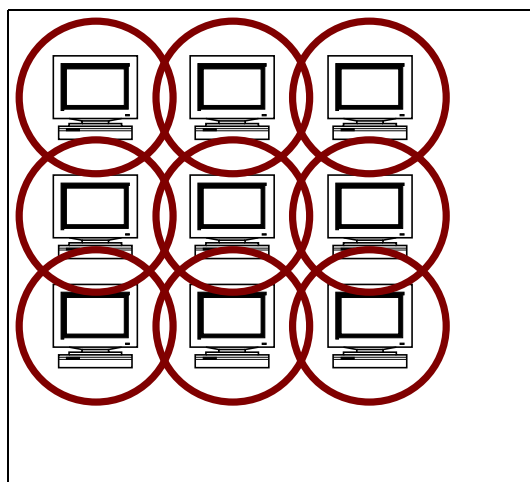



Figura 6-2.8: Topologia em Malha

6.2.9. Topologia celular

A *topologia celular* consiste em áreas circulares ou hexagonais, cada uma tendo um nó individual no centro.

A topologia celular é uma área geográfica dividida em regiões (células) para fins de tecnologia sem-fio, uma tecnologia que se torna cada vez mais importante. Não há links físicos em uma topologia celular, apenas ondas eletromagnéticas. Às vezes, os nós de recepção (por exemplo, o telefone celular de carro) se movem e, às vezes, os nós de envio se movem (por exemplo, os links de comunicação por satélites). A vantagem óbvia de uma

topologia celular (sem fio) é que não há outros meios tangíveis que não a atmosfera terrestre ou o vácuo do espaço interplanetário (e satélites). As desvantagens são que os sinais estão presentes em todos os lugares de uma célula e, assim, são suscetíveis a interferências (provocadas pelo ser humano e pelo meio ambiente) e às violações na segurança (por exemplo, o monitoramento eletrônico e roubo de serviço).

As tecnologias celulares comunicam-se umas com as outras diretamente (embora as limitações impostas pela distância e a interferência às vezes tornem essa comunicação extremamente difícil) ou comunicam-se apenas com suas células adjacentes, o que é muito ineficiente. Como regra, as topologias baseadas em celulas são integradas a outras topologias, independentemente de usarem a atmosfera ou satélites.

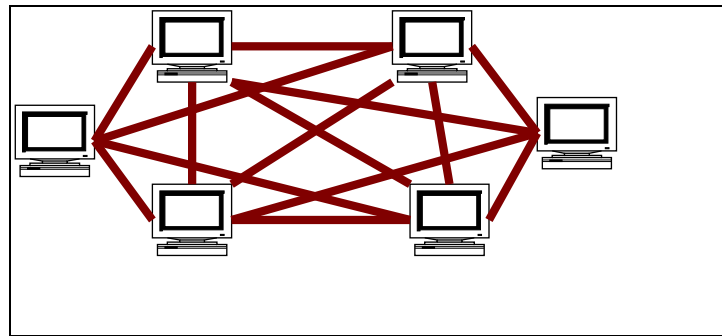



Figura 6-2.9: Topologia Celular

6.3. MODELO OSI DE ARQUITETURA

O final da década de 70 apresentava um panorama curioso em termos de comunicação de dados em redes de computadores: por um lado, uma perspectiva de crescimento vertiginoso causado pelo investimento e desenvolvimento que estavam sendo feitos, mas por outro lado uma tendência que poderia acarretar uma profunda crise no setor, a heterogeneidade de padrões entre os fabricantes, praticamente impossibilitando a interconexão entre sistemas de fabricantes distintos.

Então os fabricantes começaram a perseguir alguns objetivos necessários para a implementação de um sistema aberto. Esses objetivos são:

- interoperabilidade: capacidade que os sistemas abertos possuem de troca de informações entre eles, mesmo que sejam fornecidos por fabricantes diversos;
- interconectividade: é a maneira através da qual se pode conectar computadores de fabricantes distintos;
- portabilidade da aplicação: é a capacidade de um software de rodar em várias plataformas diferentes;
- escalabilidade: capacidade de um software rodar com uma performance aceitável em computadores de capacidades diversas, desde computadores pessoais até supercomputadores.

Para se atingir estes objetivos, a ISO (International Organization for Standardization) passou a se ocupar em criar um padrão de arquitetura aberta e baseada em camadas. Foi então definido o Modelo de Referência para Interconexão de Sistemas Abertos (Reference Model for Open Systems Interconection - RM OSI).

A utilização de um ambiente de sistema aberto nos oferece algumas vantagens, como:

- liberdade de escolha entre soluções de diversos fabricantes;
- acesso mais rápido a novas tecnologias e a preços mais acessíveis, já que é mais barato e rápido fabricar produtos baseados em uma plataforma padrão;
- redução de investimentos em novas máquinas, já que os sistemas e os softwares de aplicação são portáveis para os vários tipos de máquinas existentes.

A adoção de um modelo baseado em camadas também não é arbitrária. Considerando que uma rede de computadores tem como objetivo o processamento de tarefas distribuídas

pela rede de forma harmônica e cooperativa entre os vários processos de aplicação, o projeto desta deve levar em conta vários fatores, como:

- considerar todos os eventos possíveis de acontecer durante a comunicação;
- conhecer todos os efeitos e causas destes eventos;
- especificar em detalhes todos os aspectos técnico-operacionais dos meios físicos a serem utilizados como suporte à comunicação;
- detalhes das próprias aplicações a serem executadas.

Podemos perceber, então, que o problema é extremamente complexo e abrangente. A fim de se lidar com esta complexidade (facilitando a implementação e manutenção), projeta-se a rede como um conjunto de camadas.

Este conjunto de camadas é hierárquico, ou seja, cada camada baseia-se na camada inferior. Reduzindo-se o projeto global da rede ao projeto de cada uma das camadas, simplifica-se consideravelmente o trabalho de desenvolvimento e de manutenção. O projeto de uma camada é restrito ao contexto dessa camada e supõe que os problemas fora deste contexto já estejam devidamente resolvidos.

Na realidade existem duas vantagens práticas na utilização de uma arquitetura em camadas. Em primeiro lugar, a complexidade do esforço global de desenvolvimento é reduzida através de abstrações (não interessa para uma determinada camada como as demais implementam o fornecimento de seus serviços, só o que elas oferecem). Na arquitetura hierárquica, a camada (N) sabe apenas que existem a camada (N-1), prestadora de determinados serviços e a camada (N+1), que lhe requisita os serviços. A camada (N) não toma conhecimento da existência das camadas (N±2), (N±3), etc.

O segundo aspecto é relacionado com a independência entre as camadas. A camada (N) preocupa-se apenas em utilizar os serviços da camada (N-1), independentemente do seu protocolo. É assim que uma camada pode ser alterada sem mudar as demais (facilidade de manutenção) - desde que os serviços que ela presta não sejam modificados. É assim também que novas aplicações podem ser implementadas, na camada apropriada, aproveitando os mesmos serviços já fornecidos pelas outras camadas (redução dos esforços para evoluções).

Porém a elaboração de um sistema aberto passa por algumas etapas obrigatórias que podemos observar claramente na definição do modelo OSI, da ISO:

- definição do modelo do sistema aberto (padrão para arquitetura do sistema aberto);

• definição dos padrões dos componentes que fazem parte do modelo (padrões de interoperabilidade e portabilidade), não só os relacionados à comunicação, mas também alguns não relacionados, como estrutura de armazenamento de dados, etc;

• seleção dos perfis funcionais.

Podemos observar que o modelo OSI da ISO corresponde exatamente ao primeiro ítem citado acima. O modelo OSI é um modelo de referência e define apenas a arquitetura do sistema. O padrão criado para o modelo OSI, então, define exatamente o que cada camada deve fazer, mas não define como isto será feito, ou seja, define os serviços que cada camada deve prestar, mas não o protocolo que o realizará. Este primeiro passo já está bem definido pela ISO.

A definição dos protocolos de cada camada, então, fica por conta do segundo passo. Esta parte também está definida pela ISO, mas é realizado por grupos de estudo diversos. Este passo é uma tarefa muito dinâmica, pois novas tecnologias de transmissão surgem a todo instante. Portanto, por um lado temos alguns padrões bem documentados, mas por outro, temos tecnologias emergentes que precisam ser adaptadas às condições do modelo OSI e ainda estão em processo de definição.

Já a terceira etapa não é uma fase de responsabilidade da ISO. Esta etapa de definição de perfis funcionais é realizada por cada país, que escolhe os padrões que lhe cabem baseados em condições tecnológicas, base instalada, visão futura, etc. Por exemplo, no Brasil temos o Perfil Funcional do Governo Brasileiro. A escolha do Perfil Funcional é uma etapa importante, pois apesar de dois sistemas seguirem o Modelo OSI, se eles adotarem perfis diferentes, eles nunca vão conseguir interoperar.

A arquitetura OSI foi desenvolvida a partir de três elementos básicos:

• os processos de aplicação existentes no ambiente OSI;

• as conexões que ligam os processos de aplicação e que lhes permitem trocar informações;

• os sistemas.

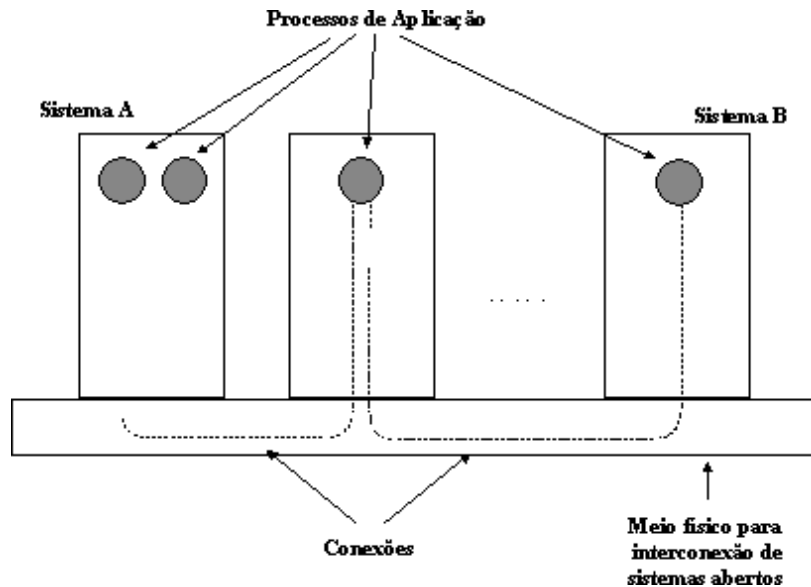



Figura 6-3.1: Processos de aplicação, conexões e sistemas

O desenho abaixo nos dá uma idéia da arquitetura de uma máquina pertencente a um sistema de comunicação:

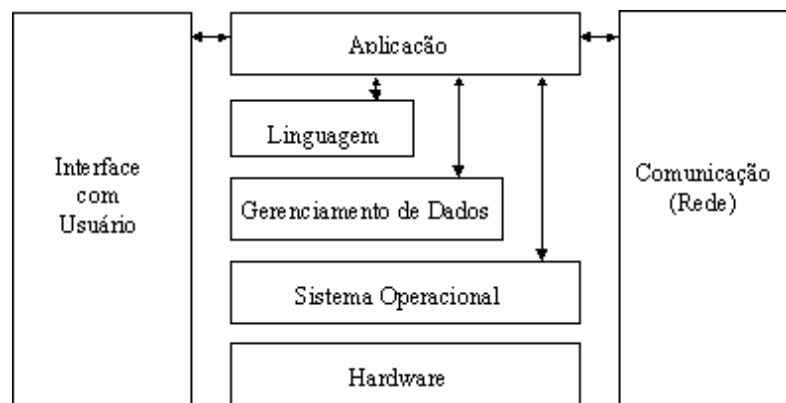



Figura 6-3.2: Arquitetura de uma máquina do sistema

Hardware: provê a infra-estrutura necessária (no nível mais baixo) para o processamento da aplicação, como a manipulação de bits, acesso a disco, etc. · Sistema operacional: provê os serviços básicos de acesso a hardware, etc. · Gerenciamento de dados: cuida de tarefas como o acesso, manipulação e troca de vários tipos de dados. A consistência nesta tarefa é um grande passo rumo à portabilidade de aplicações. Existem várias formas de implementação de acesso a bancos de dados, mas a mais comum e aceita pela indústria é a SQL (Structure Query Language). ·

Linguagem: têm sido feitos esforços em relação à criação de uma linguagem com independência da plataforma, de forma a prover a portabilidade de código. · Interface com o

usuário: um dos principais fatores de portabilidade, já que provê a interface com o usuário da aplicação. Cada vez mais estão sendo desenvolvidas interfaces gráficas e orientadas a objetos baseadas em janelas, ícones e menus.

Os principais padrões para desenvolvimento de interfaces gráficas são X Window e Motif. · Comunicação: a parte de comunicação é o objeto principal do nosso estudo. Ela vai prover a comunicação e interoperação entre máquinas e sistemas diferentes, cuidando de características como padrões de interoperação, endereçamento, etc.

O modelo OSI, então, se encaixa na figura 6-3.2 como um conjunto de funções que possibilitam que máquinas distintas possam se comunicar e trocar informações. Ele possui sete camadas (figura 6-3.3), onde cada camada é responsável por uma determinada função específica. Os princípios utilizados para se chegar a estas camadas são:

- uma camada deve ser criada onde é necessário um nível de abstração diferente;
- cada camada deve desempenhar uma função bem definida;
- a função de cada camada deve ser definida tendo em vista a definição de protocolos padrões internacionais;
- as fronteiras entre as camadas devem ser escolhidas de forma a minimizar o fluxo de informações através das interfaces;
- o número de camadas deve ser grande o suficiente para que não seja preciso agrupar funções em uma mesma camada por necessidade, e pequeno o suficiente para que a arquitetura fique manejável.

Cada camada é usuária dos serviços prestados pela camada imediatamente inferior e presta serviços para a camada imediatamente superior. Esta troca de informações entre as camadas adjacentes ocorre por meio da troca de primitivas de serviços nas interfaces entre as camadas.

Apesar do modelo OSI estar dividido em sete níveis, pode-se considerar genericamente que as três camadas mais baixas cuidam dos aspectos relacionados à transmissão propriamente dita e a camada de transporte lida com a comunicação fim-a-fim, enquanto que as três camadas superiores se preocupam com os aspectos relacionados à aplicação, já no nível de usuário.

A comunicação entre sistemas ocorre ao nível de camadas, ou seja, a camada de aplicação do sistema A se comunica com a camada de aplicação do sistema B e assim por diante até o nível físico, onde ocorre a comunicação física entre os sistemas.

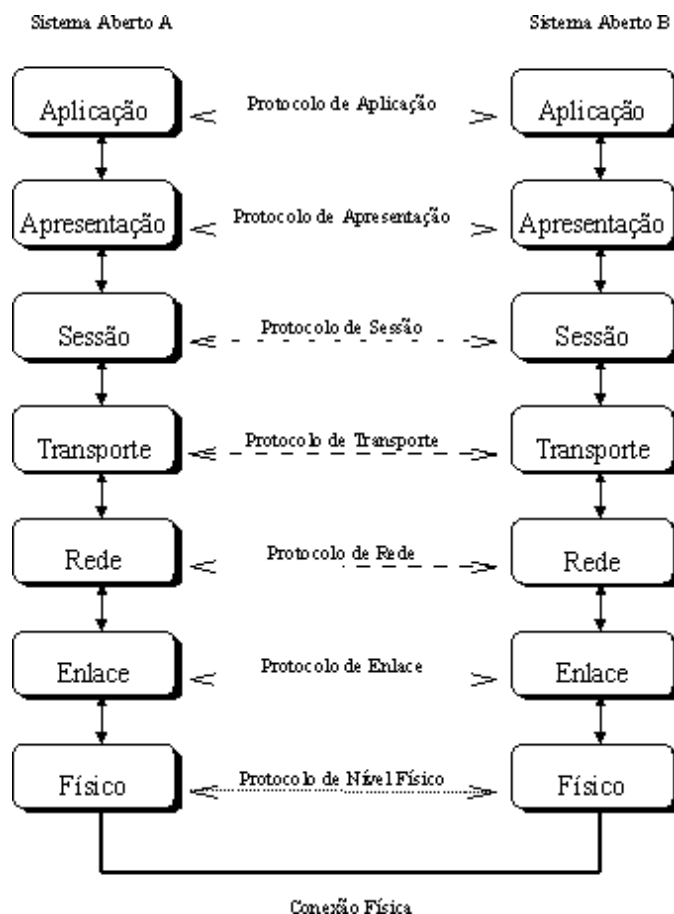



Figura 6-3.3: Modelo OSI

Uma maneira bastante fácil e simplista de se enxergar a funcionalidade de um modelo em camadas, como o modelo OSI, é imaginar que cada camada tem como função adicionar um cabeçalho aos dados do usuário a serem transmitidos para outro sistema (figura 6-3.4). Deste modo a função de cada camada do outro sistema é exatamente a inversa, ou seja, retirar os cabeçalhos dos dados que chegam e entregá-los ao usuário em sua forma original.

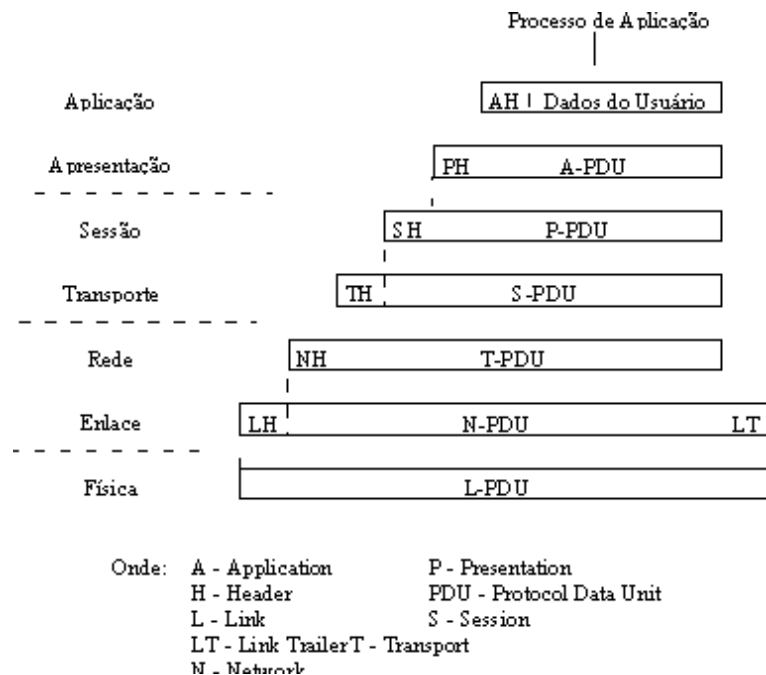



Figura 6-3.4: Transferência de Dados entre Camadas

Os dados entregues pelo usuário à camada de aplicação do sistema recebem a denominação de SDU (Service Data Unit). A camada de aplicação, então, junta à SDU (no caso, os dados do usuário) um cabeçalho chamado PCI (Protocol Control Information). O objeto resultante desta junção é chamado de PDU (Protocol Data Unit), que corresponde à unidade de dados especificada de um certo protocolo da camada em questão.

6.3.1. Primitivas de Serviços

As primitivas de serviços são informações trocadas entre duas camadas adjacentes de forma a realizar um serviço. No modelo OSI são definidos quatro tipos de primitivas:

- Pedido (Request): utilizada para solicitar ou ativar um determinado serviço;
- Indicação (Indication): informa a ocorrência de um determinado evento;
- Resposta (Response): utilizada para responder a um determinado evento;
- Confirmação (Confirmation): utilizada para confirmar a execução de um serviço solicitado.

As primitivas possuem parâmetros de entrada e saída. Por exemplo, em um pedido de conexão, os parâmetros podem especificar a máquina à qual se deseja conectar, o tipo de

serviço desejado e o tamanho máximo de mensagem a ser utilizada. Os parâmetros em uma aceitação de conexão podem conter a identidade do solicitante, o tipo de serviço e o tamanho máximo de mensagem proposto. Quem cuida dos detalhes desta negociação é o protocolo. Por exemplo, caso duas propostas para o tamanho máximo das mensagens trocadas seja conflitante, o protocolo deve decidir qual das duas será aceita.

Os serviços prestados podem ser basicamente de dois tipos: confirmado e não confirmado. No serviço confirmado, há um pedido, uma indicação, uma resposta e uma confirmação. Já no serviço não confirmado, há apenas um pedido e uma indicação. Um exemplo de um serviço confirmado é o estabelecimento de uma conexão, enquanto que a desconexão é um serviço não confirmado. Vejamos o exemplo de um serviço de conexão na figura 6-3.5.

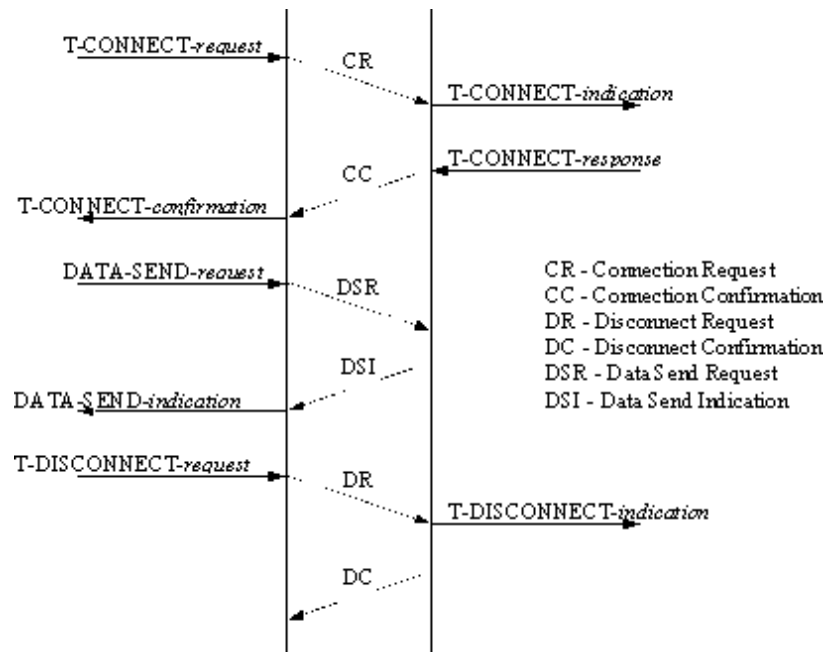



Figura 6-3.5: Diagrama de Tempo de Estabelecimento de Conexão

Este serviço pode ser descrito da seguinte forma:

- request.CONEXÃO - solicita o estabelecimento de uma conexão;
- indication.CONEXÃO - informa à parte chamada;
- response.CONEXÃO - entidade chamada aceita ou rejeita chamadas;
- confirmation.CONEXÃO - indica ao solicitante se a chamada foi aceita;
- request.DADOS - solicita a transmissão de dados;
- indication.DADOS - avisa sobre a chegada de dados;

• request.DESCONEXÃO - solicita que a conexão seja liberada;

• indication.DESCONEXÃO - informa ao parceiro sobre o pedido.

Um exemplo muito didático é a analogia com o sistema telefônico. Por exemplo, você liga para uma pessoa e a convida para sair:

• request.CONEXÃO - você disca o telefone da pessoa;

• indication.CONEXÃO - o telefone dela toca;

• response.CONEXÃO - ela atende o telefone;

• confirmation.CONEXÃO - você ouve o sinal de chamada parar de tocar;

• request.DADOS - você convida a pessoa para sair;

• indication.DADOS - ela ouve seu convite;

• request.DADOS - ela responde que sim;

• indication.DADOS - você ouve a aceitação dela;

• request.DESCONEXÃO - você desliga o telefone;

• indication.DESCONEXÃO - ela ouve e desliga também.

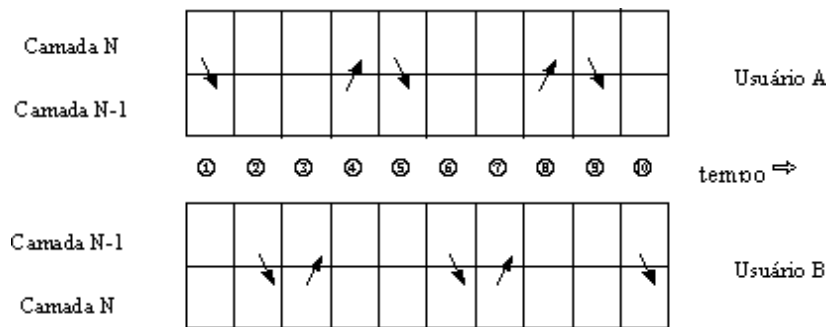



Figura 6-3.6: Diagrama de Tempo da Conversação Telefônica

A camada N é o usuário, ou seja, você e a pessoa com quem está falando. A camada N-1 é a operadora do serviço. De um modo bastante simplificado, este exemplo nos mostra a troca de primitivas em uma conversação genérica, mas que pode ser perfeitamente aplicada a situações mais complexas, como o modelo OSI.

6.3.2. Serviços e Protocolos

Faz-se necessário neste ponto deixar bem clara a distinção entre serviços e protocolos. Um serviço é um conjunto de primitivas que uma camada oferece à camada superior adjacente, ou seja, é uma interface entre duas camadas onde a inferior se comporta como provedora do serviço e a superior a usuária do serviço. O serviço define as operações que a camada está preparada para realizar em nome de seus usuários, mas não diz nada a respeito do modo como isso deve ser implementado.

Já um protocolo é um conjunto de regras que governa o formato e significado dos quadros, pacotes ou mensagens trocados entre entidades parceiras dentro de uma mesma camada. Os protocolos são utilizados para implementar os serviços, não sendo diretamente visíveis aos usuários, ou seja, o protocolo utilizado pode ser modificado, desde que o serviço oferecido ao usuário permaneça o mesmo.

Devemos sempre lembrar que ao se falar em serviços, estamos falando de camadas adjacentes (níveis diferentes, no mesmo sistema), e ao se falar em protocolos falamos de entidades pares (no mesmo nível, em sistemas diferentes).

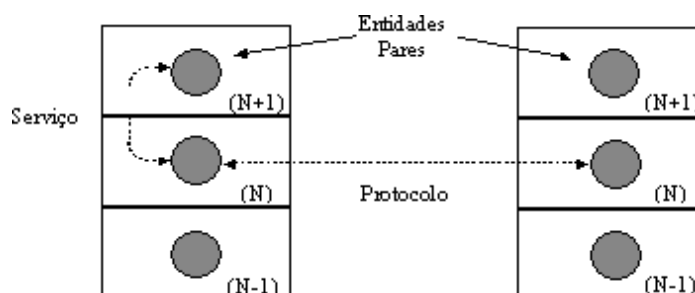



Figura 6-3.7: Serviços e protocolos no modelo OSI

Os serviços providos pela camada (N) são disponíveis para a entidade (N+1) através dos SAP´s (Service Access Point). Os SAP´s são interfaces lógicas entre as entidades (N) e (N+1). Portanto, quando a entidade (N+1) precisa utilizar o serviço provido pela camada (N), ela o busca no SAP(N).

As informações entre entidades (N+1) são trocadas através de uma associação chamada conexão (N) , estabelecida na camada (N) usando o protocolo (N). A figura 6-3.8 ilustra este conceito.

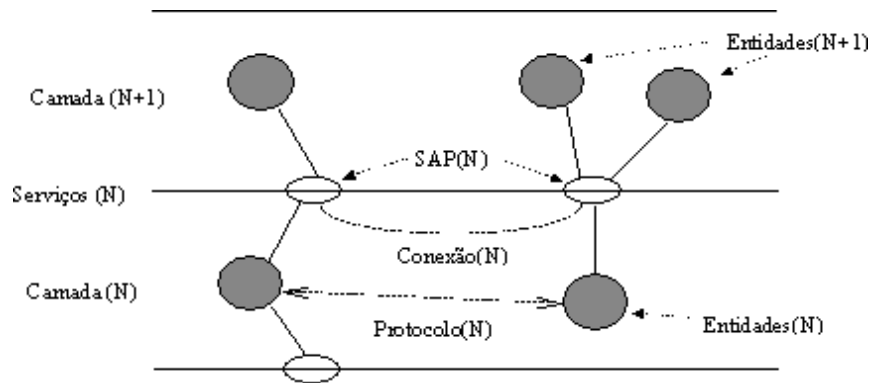



Figura 6-3.8: SAP´s e conexões

6.3.3. Camada Física

A camada física é a única camada que possui acesso ao meio de transmissão da rede devendo, portanto, se preocupar com fatores como as especificações elétricas, mecânicas, funcionais e procedurais da interface física entre o equipamento e o meio de transmissão. Ou seja, a camada física tem como função básica a adaptação do sinal ao meio de transmissão atendendo as seguintes caracteísticas:

- mecânicas: propriedades físicas da interface com o meio físico de transmissão, incluindo, por exemplo, o tipo de conector utilizado;
- elétricas: se relacionam com a representação de um bit em termos de, por exemplo, nível de tensão utilizado e taxa de transmissão de bits;
- funcionais: definem as funções a serem implementadas por esta interface;
- procedurais: especificam a seqüência de eventos trocados durante a transmissão de uma série de bits através do meio de transmissão.

A camada física possui as seguintes funções:

- estabelecimento/encerramento de conexões: ativa e desativa conexões físicas mediante a solicitação de entidades da camada de enlace;
- transferência de dados: a unidade de transmissão utilizada é o bit. O nível físico tem como função transmitir os bits na mesma ordem em que chegam da camada de enlace (no sistema de origem) e entregá-los à camada de enlace na mesma ordem que chegaram (no sistema de destino);
- gerenciamento das conexões: gerência da qualidade de serviço das conexões físicas estabelecidas. Deve monitorar taxa de erros, disponibilidade de serviço, taxa de transmissão, atraso de trânsito etc.

Os padrões de nível físico utilizados são, por exemplo, X.21, X.21 bis, V.24, V.28, RS-232, I.430, I.431 etc.

6.3.4. Camada de Enlace

A camada de enlace tem o objetivo de prover uma conexão confiável sobre um meio físico. Sua função básica é detectar e, opcionalmente, corrigir erros que por ventura ocorram no nível físico.

As suas principais funções são:

- estabelecimento e liberação da conexão de enlace sobre conexões físicas ativas;
- *splitting* da conexão de enlace: desta forma pode haver uma conexão de enlace sobre várias conexões físicas;
- montagem e delimitação de quadros (*framing*): montagem de quadros a partir de unidades de quadros de serviços provindas da camada de rede e reconhecimento de quadros a partir da cadeia de bits vinda do nível físico;
- controle de fluxo: controla a taxa de transmissão dos quadros, evitando que o sistema transmissor envie dados a uma taxa maior do que o receptor consegue processar. Utiliza para isso mecanismos como *stop-and-wait*, *positive acknowledgment* (ACK) e *sliding window*.
- controle de acesso: gerência do acesso ao meio de transmissão;
- controle de erro: a camada de enlace deve detectar erros de transmissão, de formato e de operação devidos a problemas de conexão física ou mau funcionamento da própria camada. Os erros mais comumente detectados são erros devidos a perdas, duplicação, não-ordenação e danificação de quadros.
- controle de seqüência: as unidades de dados de serviço de enlace devem ser entregues à entidade de rede de destino na mesma ordem em que são recebidas da entidade de rede de origem;
- gerenciamento: a camada de enlace deve exercer algumas funções de gerenciamento relacionadas à qualidade do serviço prestado, caracterizada por: tempo médio entre erros irrecuperáveis, taxa de erro residual decorrente da alteração, perda, duplicação e não-ordenação dos quadros, disponibilidade do serviço, atraso de trânsito e throughput (vazão).

O protocolo de enlace mais conhecido é o HDLC, antecessor de outros como LAPB, LAPD e LLC.

6.3.5. Camada de Rede

A camada de rede deve tornar transparente para a camada de transporte a forma como os recursos dos níveis inferiores são utilizados para implementar conexões de rede. Deve também equalizar as diferenças entre as diversas sub-redes utilizadas de modo a fornecer um serviço único a seus usuários (independentemente da rede utilizada).

Suas principais funções são:

- roteamento: determinação das rotas apropriadas para a transmissão dos dados entre dois endereços (origem e destino) através de algoritmos de roteamento;
- multiplexação da conexão de rede: várias conexões de rede podem ser multiplexadas sobre uma única conexão de enlace, a fim de otimizar a utilização desta última;
- segmentação e blocagem: caso as sub-redes envolvidas em uma comunicação fim-a-fim possuam diversos tipos e tamanhos de quadros, a camada de rede deve exercer funções de segmentação de quadros e remontagem destes no destino;
- controle de erro: detecta e, dependendo da qualidade do serviço exigida, até corrige erros de alteração, perda, duplicação e não-ordenação das unidades de dados;
- seqüenciação: a camada de rede é responsável por manter a ordem das unidades de dados de serviço de rede a serem transmitidas na rede e recebidas pela camada de transporte no destino;
- controle de fluxo: controle da taxa em que os dados são transmitidos, de forma que o transmissor não envie mais dados do que o receptor tenha capacidade de receber;
- transferência de dados expressos: a transmissão de dados expressos tem por finalidade estabelecer prioridade de transmissão para alguns dados (como sinalização e interrupção) sobre os dados normais;
- seleção de serviço: permite a escolha do serviço de rede, de modo a garantir que os serviços oferecidos pelas diversas sub-redes sejam equivalentes;
- gerenciamento: a camada de rede deve efetuar tarefas de gerenciamento relacionadas à qualidade de serviço oferecida, caracterizada pelos parâmetros citados acima.

A camada de rede pode prestar serviços orientados à conexão (CONS - Connection Oriented Network Service) ou serviços não-orientados à conexão (CLNS - ConnectionLess Oriented Network Service). Um exemplo de protocolo utilizado na camada de rede é o X.25.

6.3.6. Camada de Transporte

A camada de transporte provê mecanismos que possibilitam a troca de dados fim-a-fim, ou seja, a camada de transporte não se comunica com máquinas intermediárias na rede, como pode ocorrer com as camadas inferiores.

As principais funções da camada de transporte são:

- estabelecimento e liberação da conexão de transporte: para se estabelecer a conexão, devem ser negociadas a classe de protocolo a ser utilizada, o tamanho máximo das unidades de dados de protocolo, a utilização ou não do serviço de transferência de dados expressos, parâmetros de qualidade de serviço (throughput, atraso de trânsito, prioridade, taxa de erro residual, etc) etc;
- controle de seqüência e controle de erro: numeração e reconhecimento explícito dos dados a fim de evitar perdas, duplicação ou entrega fora de ordem;
- controle de fluxo: a técnica de controle de fluxo utilizada na camada de transporte é a técnica de alocação de crédito, parecida com a janela deslizante;
- segmentação, blocagem e concatenação fim-a-fim: adapta o tamanho da unidade de dados utilizada para as camadas de transmissão;
- monitoração da qualidade do serviço: a monitoração da qualidade de serviço deve ser constante, caso contrário deve ser gerada uma notificação à camada de sessão;
- transferência de dados expressos: a camada de transporte deve possibilitar esta funcionalidade;
- gerenciamento: funções de gerenciamento relacionadas à qualidade de serviço prestado às entidades de sessão.

6.3.7. Camada de Sessão

A camada de sessão é a responsável pelo estabelecimento de sessões entre dois usuários permitindo o transporte ordinário de dados (assim como a camada de transporte), porém com alguns serviços mais refinados, que podem ser úteis em algumas aplicações.

Alguns serviços que a camada de sessão deve prover são, por exemplo:

- gerência do controle de diálogo: a troca de informações entre entidades em um circuito half-duplex deve ser controlada através da utilização de tokens. A camada de sessão é responsável pela posse e entrega destes tokens, ajudando a controlar de quem é a vez de transmitir;
- sincronização: para se evitar, por exemplo, a perda de um volume de dados muito grande que estiver sendo transmitido em uma rede não confiável, utiliza-se o conceito de ponto de sincronização. O ponto de sincronização corresponde a marcas lógicas posicionadas ao longo do diálogo. Toda vez que um usuário recebe um ponto de sincronização, deve enviar uma resposta, confirmando que este foi recebido. Caso a transmissão, por algum motivo, seja interrompida, ela pode ser reiniciada a partir do último ponto de sincronização confirmado;
- gerenciamento da camada de sessão.

6.3.8. Camada de Apresentação

A camada de apresentação, ao contrário das camadas inferiores, já não se preocupa com os dados a nível de bits, mas sim com a sua sintaxe, ou seja, sua representação. Nela é definida a sintaxe abstrata, ou seja, a forma como os tipos e os valores dos dados são definidos, independentemente do sistema computacional utilizado e a sintaxe de transferência, ou seja, a maneira como é realizada esta codificação. Por exemplo, através da sintaxe abstrata define-se que um caracter A deve ser transmitido. A sintaxe de transferência especifica, então, como este dado será codificado em ASCII ou EBCDIC ao ser entregue à camada de sessão.

Outras funções que a camada de apresentação pode executar são a criptografia e compressão de dados.

6.3.9. Camada de Aplicação

Basicamente, as funções da camada de aplicação são aquelas necessárias à adaptação dos processos de aplicação ao ambiente de comunicação. A camada de aplicação é estruturada modularmente para permitir a flexibilidade das funções e das formas, para se determinar os requisitos de comunicação de cada aplicação distribuída. A camada de aplicação deve seguir o que é determinado na norma ISO 9545. Ela é formada por várias ASE's (Elemento de Serviço de Aplicação), que são os componentes básicos das AE's (Entidade de Aplicação). Uma AE é a função que um processo de aplicação utiliza para se comunicar com os seus pares. Um processo de aplicação pode utilizar diversas AE's, cada uma das quais provendo um conjunto de definições de cada uma das funções e das regras que governam o uso destas funções.

O componente básico de uma AE é um Elemento de Serviço de Aplicação (ASE). Um ASE é um elemento que define uma função ou um conjunto de funções que ajudam na realização da aplicação. Desta forma, pode-se imaginar que um AE é um grande programa feito de muitos sub-programas ou procedimentos, que são os ASE.

As principais funções da camada de aplicação são:

- Seleção do modo de diálogo (full duplex ou half duplex);
- Determinação da qualidade de serviço aceitável na conexão: retardo de transferência, taxa de erro tolerável, etc;
- Identificação dos futuros parceiros na comunicação: por nome ou endereço;
- Especificação de aspectos relativos à segurança: controle de acesso, integridade de dados, etc.

6.4. MODELO TCP/IP

Embora o modelo de referência OSI seja universalmente reconhecido, o padrão aberto técnico e histórico da Internet é o *Transmission Control Protocol/Internet Protocol (TCP/IP).* O *modelo de referência TCP/IP* e a *pilha de protocolos TCP/IP* tornam possível a comunicação de dados entre dois computadores quaisquer, em qualquer parte do mundo, praticamente à velocidade da luz. O modelo TCP/IP tem importância histórica, assim como os padrões que permitiram que as indústrias de telefonia, energia elétrica, estradas de ferro e videoteipe se desenvolvessem.

O Departamento de Defesa dos Estados Unidos (*DoD*) desenvolveu o modelo de referência TCP/IP porque queria uma rede que pudesse sobreviver a qualquer condição, mesmo a uma guerra nuclear. Para ilustrar melhor, imagine um mundo em guerra, entrecruzado por diferentes tipos de conexões: cabos, microondas, fibras óticas e links de satélites. Imagine, então, que você precise que informações/dados (na forma de pacotes) trafeguem, independentemente da condição de qualquer nó ou rede particular na internetwork (que, nesse caso, pode ter sido destruída pela guerra). O Departamento de Defesa dos Estados Unidos quer que seus pacotes cheguem, todas as vezes, em qualquer condição, de um ponto a qualquer outro. Foi esse complexo problema de projeto que levou à criação do modelo TCP/IP e que se tornou, desde então, o padrão com o qual a Internet se desenvolveu.

Quando ler sobre as camadas do modelo TCP/IP, tenha em mente o objetivo inicial da Internet; isso vai ajudar a entender porque certas coisas são como são. O modelo TCP/IP tem quatro camadas: a camada de aplicação, a camada de transporte, a *camada de Internet* e a camada de acesso à rede. É importante notar que algumas das camadas do modelo TCP/IP têm o mesmo nome das camadas no modelo OSI. Não confunda as camadas dos dois modelos, porque a camada de aplicação tem funções diferentes em cada modelo.

6.4.1. Camada de aplicação

Os projetistas do TCP/IP decidiram que os protocolos de mais alto nível deviam incluir os detalhes da camada de apresentação e de sessão. Eles simplesmente criaram uma

camada de aplicação que trata de protocolos de alto nível, questões de representação, codificação e controle de diálogo. O TCP/IP combina todas as questões relacionadas a aplicações em uma camada e garante que esses dados estejam empacotados corretamente para a próxima camada.

6.4.2. Camada de transporte

A camada de transporte lida com questões de qualidade de serviços, de confiabilidade, controle de fluxo e correção de erros. Um de seus protocolos, o Transmission Control Protocol (TCP), fornece formas excelentes e flexíveis de se desenvolverem comunicações de rede confiáveis com baixa taxa de erros e bom fluxo. O TCP é um protocolo orientado para conexões. Ele mantém um diálogo entre a origem e o destino enquanto empacota informações da camada de aplicação em unidades chamadas segmentos. Orientado para conexões não significa que exista um circuito entre os computadores que se comunicam (o que seria feito por uma comutação de circuitos). Significa que segmentos da camada 4 trafegam entre dois hosts para confirmar que a conexão existe logicamente durante um certo período. Isso é conhecido como comutação de pacotes.

6.4.3. Camada de Internet

A finalidade da *camada de Internet* é enviar pacotes da origem de qualquer rede na internetwork e fazê-los chegar ao destino, independentemente do caminho e das redes que usem para chegar lá. O protocolo específico que governa essa camada é chamado Internet protocol (IP). A determinação do melhor caminho e a comutação de pacotes acontecem nessa camada. Pense nisso em termos do sistema postal. Quando você envia uma carta, você não sabe como ela vai chegar ao seu destino (existem várias rotas possíveis), mas, o que realmente importa, é que ela chegue.

6.4.4. Camada de acesso à rede

O significado do nome dessa camada é muito amplo e um pouco confuso. É também chamada de camada host-rede. É' a camada que se relaciona a tudo aquilo que um pacote IP

necessita para realmente estabelecer um link físico e depois, se necessário, estabelecer outro link físico. Isso inclui detalhes de tecnologia de LAN e WAN e todos os serviços das camadas física e de enlace do OSI.

Figura 6-4.1: Modelo em camadas TCP/IP

6.4.5. Protocolos TCP/IP

O diagrama mostrado na figura é chamado *gráfico do protocolo*. Ele ilustra alguns dos protocolos comuns especificados pelo modelo de referência TCP/IP. Na camada de aplicação, você vai ver diferentes tarefas de rede que talvez não reconheça, mas que, como usuário da Internet, provavelmente usa todos os dias. Esses aplicativos incluem:

- *FTP* - File Transfer Protocol
- *HTTP* - Hypertext Transfer Protocol
- *SMTP* - Simple Mail Transfer Protocol
- *DNS* - Sistema de Nomes de Domínio
- *TFTP* - Trivial File Transfer Protocol

O modelo TCP/IP enfatiza a máxima flexibilidade, na camada de aplicação, para desenvolvedores de software. A camada de transporte envolve dois protocolos - transmission control protocol (TCP) e *user datagram protocol (UDP)*. A camada mais baixa, a camada de acesso à rede, refere-se à tecnologia de LAN ou WAN específica que está sendo usada.

É importante indicar a diferença destes dois protocolos e os serviços a que eles se propõem. O TCP é um protocolo que prove um serviço orientado a conexão, isto é, com garantia de entrega de pacotes, que espera a confirmação de cada datagrama e os restransmite caso estes sejam perdidos. Por isso ele é mais lento, se comparado ao UDP, e é utilizado em

aplicações onde a perda de pacotes é mais sensível. Já o UDP é um protocolo que provê um serviço não orientado a conexão, isto é, sem garantida de entrega de pacotes e sem retransmissão, caso algum seja perdido. Sendo assim, ele é mais rápido e utilizado em aplicações em tempo real, onde a velocidade é prioritária e a perda de alguns pacotes não é tão sentida. Nessas situações, a menor sobrecarga do UDP permite serviços de broadcast.

UDP	**TCP**
Serviços sem conexão; nenhuma sessão é estabelecida entre os hosts.	Serviço orientado por conexão; uma sessão é estabelecida entre os hosts.
UDP não garante ou confirma entrega ou seqüência de dados.	TCP garante a entrega através do uso de confirmações e entrega seqüenciada dos dados.
Os programas que usam UDP são responsáveis por oferecer a confiabilidade necessária ao transporte de dados.	Os programas que usam TCP têm garantia de transporte confiável de dados.
UDP é rápido, necessita de baixa sobrecarga e pode oferecer suporte à comunicação ponto-a-ponto e ponto-a-multi-ponto.	TCP é mais lento, necessita de maior sobrecarga e pode oferecer suporte apenas à comnicação ponto-a-ponto.

No modelo TCP/IP, não importa que aplicativo solicite serviços de rede, nem que protocolo de transporte esteja sendo usado, haverá apenas um protocolo de rede, o internet protocol, ou IP, por "baixo" dele. Isso é uma decisão deliberada de projeto. O *IP* serve como um protocolo universal que permite que qualquer computador, em qualquer lugar, se comunique a qualquer momento.

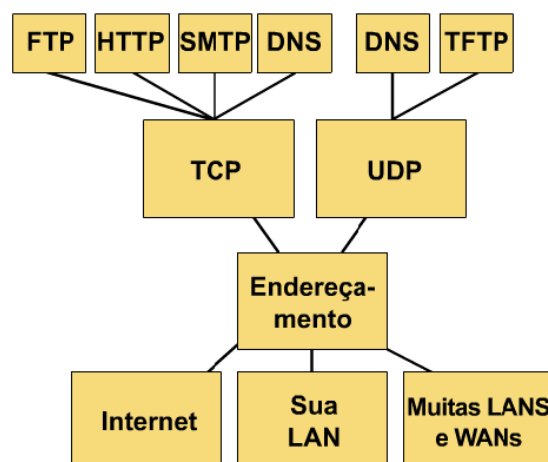



Figura 6-4.2: Protocolos TCP/IP

6.4.6. Comparando o modelo TCP/IP e o modelo OSI

Se compararmos o modelo OSI e o modelo TCP/IP, iremos notar que eles têm semelhanças e diferenças. Por exemplo:

Semelhanças

- Ambos têm camadas
- Ambos têm camadas de aplicação, embora incluam serviços muito diferentes
- Ambos têm camadas de transporte e de rede comparáveis
- A tecnologia de comutação de pacotes (e não comutação de circuitos) é presumida por ambos
- Os profissionais de rede precisam conhecer ambos

Diferenças

- O TCP/IP combina os aspectos das camadas de apresentação e de sessão dentro da sua camada de aplicação
- O TCP/IP combina as camadas física e de enlace do OSI em uma camada
- O TCP/IP parece ser mais simples por ter menos camadas
- Os protocolos TCP/IP são os padrões em torno dos quais a Internet se desenvolveu, portanto o modelo TCP/IP ganha credibilidade por causa do intenso uso dos seus protocolos. Pelo contrário geralmente as redes não são desenvolvidas de acordo com o modelo OSI, embora ele seja usado como um guia.

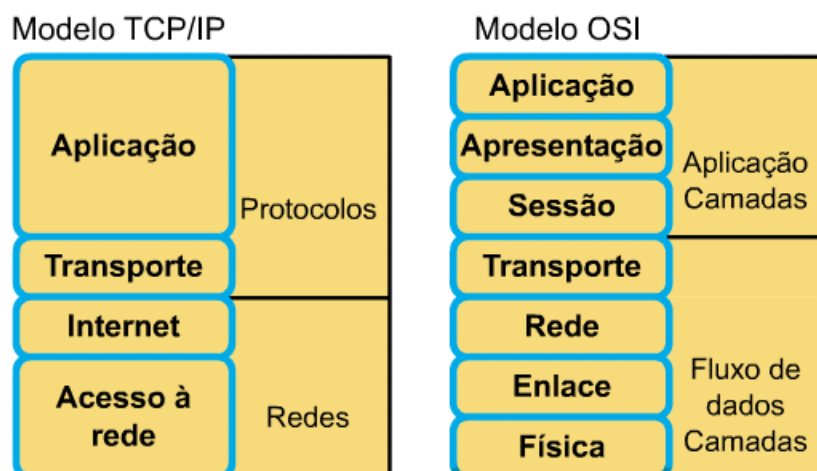



Figura 6-4.3: Comparação entre os modelos OSI e TCP/IP

6.4.7. Endereçamento IP

Um endereço IP é representado por um número binário de 32 bits. Os endereços IP são expressos como números decimais com pontos: divide-se os 32 bits do endereço em quatro *octetos* (um octeto é um grupo de 8 bits). O valor decimal máximo de cada octeto é 255.

O número de rede de um endereço IP identifica a rede à qual um dispositivo está conectado. A parte do host de um endereço IP identifica o dispositivo específico nessa rede. Como os endereços IP consistem em quatro octetos separados por pontos, um, dois ou três desses octetos podem ser usados para identificar o número de rede. De forma semelhante, até três desses octetos podem ser usados para identificar a parte do host de um endereço IP.

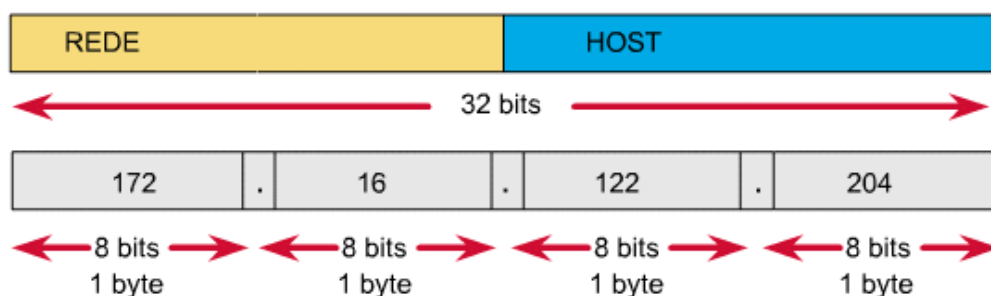



Figura 6-4.4: Exemplo de endereço IP e a divisão de identificação de rede e de host

Existem três classes de endereços IP que uma organização pode receber do American Registry for Internet Numbers (ARIN) (ou do ISP da organização). Elas são classe A, B e C. O ARIN reserva, agora, os endereços de classe A para governos por todo o mundo (embora algumas grandes empresas, como, por exemplo, a Hewlett Packard, tenham recebido um no passado) e de classe B para empresas de médio porte. A todos os outros requerentes são atribuídos endereços de classe C.

Classe A

Quando escrito em formato binário, o primeiro bit (mais à esquerda) de um endereço da classe A é sempre 0. Um exemplo de um endereço IP de classe A é 124.95.44.15. O primeiro octeto, 124, identifica o número de rede atribuído pelo ARIN. Os administradores internos da rede atribuem os 24 bits restantes. Um modo fácil de reconhecer se um dispositivo é parte de uma rede de classe A é olhar o primeiro octeto do seu endereço IP, que variará de 0 a 126. (127 *na verdade* começa com um bit 0 mas, foi reservado para propósitos especiais.)

Todos os endereços IP de classe A usam apenas os oito primeiros bits para identificar a parte da rede do endereço. Os três octetos restantes podem ser usados para a parte do host do endereço. Todas as redes que usam um endereço IP de classe A podem ter atribuídos a ela até 2 elevado a 24 (2^{24}) (menos 2), ou seja, 16.777.214 endereços IP possíveis para os dispositivos conectados à rede.

Classe B

Os dois primeiros bits de um endereço de classe B são sempre 10 (um e zero). Um exemplo de um endereço IP de classe B seria 151.10.13.28. Os dois primeiros octetos identificam o número de rede atribuído pelo ARIN. Os administradores internos da rede atribuem os 16 bits restantes. Um modo fácil de reconhecer se um dispositivo é parte de uma rede de classe B é olhar o primeiro octeto do seu endereço IP. Os endereços IP de classe B sempre têm valores variando de 128 a 191 no primeiro octeto.

Todos os endereços IP de classe B usam os primeiros 16 bits para identificar a parte da rede no endereço. Os dois octetos restantes do endereço IP podem ser usados para a parte do host do endereço. Todas as redes que usam um endereço IP de classe B podem ter atribuídos a ela até 2 elevado a 16 (2^{16}) (menos 2 novamente!), ou seja, 65.534 endereços IP possíveis para os dispositivos conectados à rede.

Classe C

Os três primeiros bits de um endereço de classe C são sempre 110 (um, um e zero). Um exemplo de um endereço IP de classe C seria 201.110.213.28. Os três primeiros octetos identificam o número de rede atribuído pelo ARIN. Os administradores internos da rede atribuem os 8 bits restantes. Um modo fácil de reconhecer se um dispositivo é parte de uma rede de classe C é olhar o primeiro octeto do seu endereço IP. Os endereços IP de classe C sempre têm valores variando de 192 a 223 no primeiro octeto.

Todos os endereços IP de classe C usam os primeiros 24 bits para identificar a parte da rede no endereço. Apenas o último octeto de um endereço IP de classe C pode ser usado para a parte do host do endereço. Todas as redes que usam um endereço IP de classe C podem ter atribuídos a ela até 2^8 (menos 2), ou seja, 254 endereços IP possíveis para os dispositivos conectados à rede.

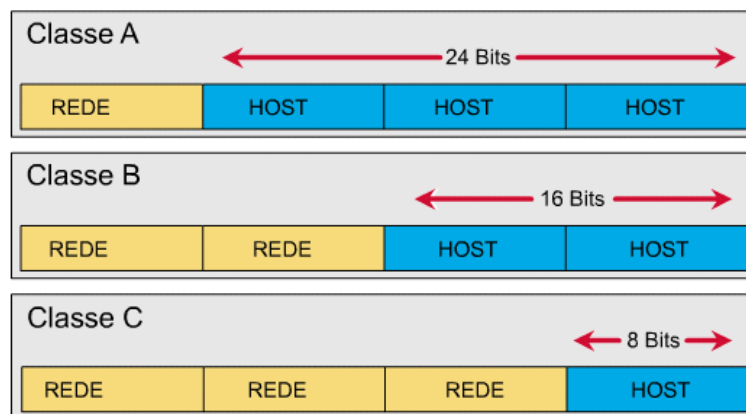



Figura 6-4.5: Classes de endereços IP

Os endereços IP identificam um dispositivo em uma rede e a rede à qual ele está ligado. Para torná-los fáceis de serem lembrados, os endereços IP são normalmente escritos em notação decimal com ponto. Portanto, os endereços IP são 4 números decimais separados por pontos. Um exemplo disso é o endereço 166.122.23.130. Tenha em mente que um número decimal é um número de base 10, o tipo de representação que usamos no dia-a-dia.

Endereço de Broadcast e de Rede

Se o seu computador quisesse se comunicar com todos os dispositivos em uma rede, seria muito pouco prático escrever os endereços IP de todos os dispositivos. Você pode tentar usar dois endereços ligados por um hífen, indicando que está se referindo a todos os dispositivos dentro de um intervalo de números, mas isso também seria pouco prático. Existe, entretanto, um método mais rápido.

Um endereço IP que termine com 0's (zeros) binários em todos os bits de host é reservado para o endereço de rede (algumas vezes chamado de endereço de *cabo*). Assim, em um exemplo de rede de classe A, 113.0.0.0 é o endereço IP da rede que contém o host 113.1.2.3. Um roteador usa um endereço IP de uma rede ao encaminhar dados na Internet. Em um exemplo de rede de classe B, o endereço IP 176.10.0.0 é o endereço de uma rede.

Os números decimais que preenchem os dois primeiros octetos em um endereço de rede de classe B são atribuídos e são números de rede. Os dois últimos octetos contêm 0's, porque esses 16 bits são números de host, e são usados para os dispositivos conectados à rede. O endereço IP do exemplo (176.10.0.0) é reservado para o endereço da rede. Ele nunca vai ser usado como um endereço para qualquer dispositivo que esteja ligado a ela.

Se você quisesse enviar dados a todos os dispositivos em uma rede, você precisaria usar o *endereço de broadcast*. Um broadcast acontece quando uma origem envia dados a todos os dispositivos em uma rede. Para assegurar que todos os dispositivos na rede vão perceber esse broadcast, a origem deve usar um endereço IP de destino que todos eles possam reconhecer e recolher. Os endereços IP de broadcast terminam com 1s binários na parte do host do endereço (*campo do host*).

Para a rede do exemplo (176.10.0.0), onde os últimos 16 bits formam o campo do host (ou parte do host do endereço), o broadcast que seria enviado a todos os dispositivos na rede incluiria um endereço de destino 176.10.255.255 (já que 255 é o valor decimal de um octeto que contém 11111111).

É importante entender o significado da parte de rede de um endereço IP, a *ID de rede*. Os hosts em uma rede podem apenas se comunicar diretamente com os dispositivos que tenham a mesma ID de rede. Pode ser que eles compartilhem o mesmo segmento físico, mas se eles tiverem números de redes diferentes, eles normalmente não poderão se comunicar um com o outro, a menos que haja outro dispositivo que possa fazer uma conexão entre as redes.

Endereçamento de Sub-redes

Os administradores de rede às vezes precisam dividir as redes, especialmente as grandes, em redes menores. Essas divisões menores são chamadas de *sub-redes* e fornecem flexibilidade de endereçamento. Os endereços de sub-rede são atribuídos localmente, normalmente pelo administrador da rede, de forma semelhante à parte do número do host dos endereços de classe A, B ou C. Além disso, como os outros endereços IP, todos os endereços de sub-rede são exclusivos.

Os endereços de sub-rede incluem a parte da rede de classe A, classe B ou classe C, mais um campo de sub-rede e um campo de host. O campo da sub-rede e o campo do host são criados a partir da parte original do host para toda a rede. A habilidade de decidir como dividir a parte original do host em novas sub-redes e campos de host permite que haja flexibilidade no endereçamento para o administrador da rede.

Para criar um endereço de sub-rede, um administrador de rede toma emprestados bits do campo do host e os designa como o campo da sub-rede. O número mínimo de bits que podem ser emprestados é 2. Se você tomasse emprestado apenas 1 bit para criar uma sub-rede, teria apenas um número de rede (a rede .0) e o número de broadcast (a rede .1). O máximo de bits que podem ser emprestados é qualquer número de bits que deixe pelo menos 2 bits para o número do host. Neste exemplo de um endereço IP de classe C, foram tomados emprestados bits do campo do host para o campo da sub-rede.

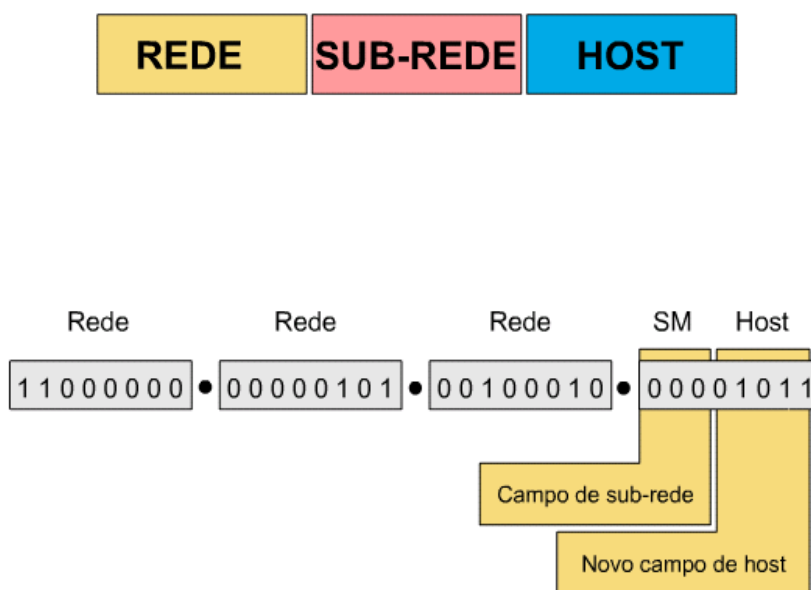



Figura 6-4.6: Exemplo de endereçamento usando sub-redes

6.4.8. Máscara de Sub-rede

A máscara de sub-rede (termo formal: prefixo de rede estendida) não é um endereço, mas determina que parte de um endereço IP é o campo de rede e que parte é o campo de host. Uma máscara de sub-rede tem o tamanho de 32 bits e tem 4 octetos, da mesma forma que um endereço IP. Para determinar a máscara de sub-rede do endereço IP de uma sub-rede específica, siga estas etapas: (1) Expresse o endereço IP da sub-rede na forma binária. (2) Substitua a parte da rede e da sub-rede do endereço composto somente por 1s. (3) Substitua a parte do host do endereço somente por 0s. (4) Como última etapa, converta a expressão binária novamente na notação decimal com ponto.

Observação: O prefixo de rede estendida inclui o número de rede de classe A, B ou C, mais o campo de sub-rede (ou número de sub-rede) que está sendo usado para estender as informações de roteamento (que, caso contrário, é apenas o número de rede)

	Rede	Subrede	Host
130.5.0.0	10000010 00000101	00000000	00000000
255.255.255.0	11111111 11111111	11111111	00000000
	Extended Network Prefix (Máscara de subrede)		

Figura 6-4.7: Exemplo de máscara de sub-rede

6.5. ATM

O ATM (*Asynchronous Transfer Mode*) é uma tecnologia de rede em nível de enlace e físico do modelo OSI, baseada na transmissão de pequenas unidades de informação de tamanho fixo e formato padronizado denominadas *células*. Foi desenvolvida para ser uma única rede de transporte para diversos serviços, tanto para redes locais como de longa distância. Assim, as redes de telefonia, televisão a cabo e comunicação de dados, hoje implementadas separadamente, poderiam ser unificadas através do ATM.

Para unificar os diversos tipos de serviços, ele exige uma camada de adaptação (AAL – *ATM Adaptation Layer*), que se localiza na parte superior do nível de enlace e efetua a adaptação dos diversos tipos de tráfego que os serviços necessitam. A figura a seguir ilustra o que foi dito acima.

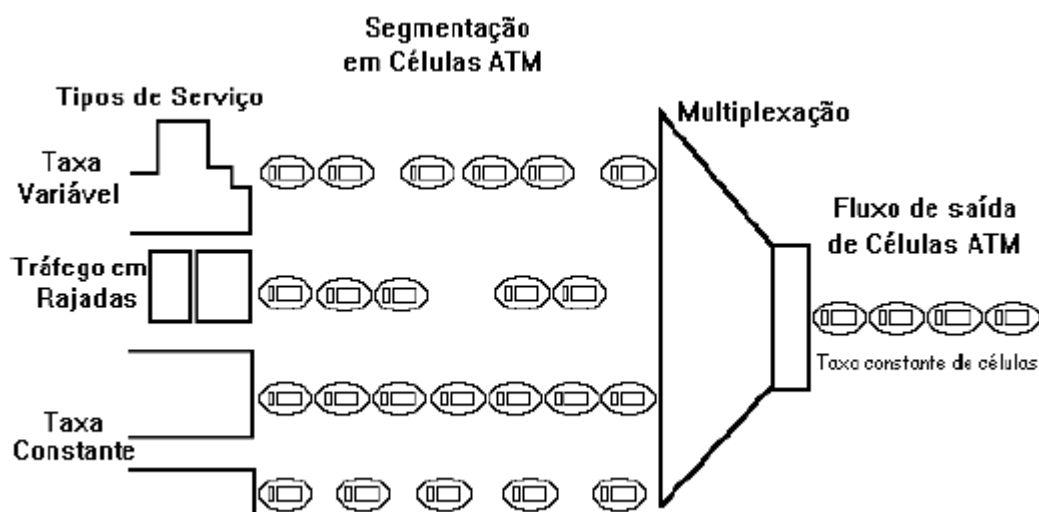



Figura 6-5.1: ATM

Serviços de multimídia com voz e imagem suportam pequenas perdas nos dados, uma vez que a perda de poucas células é imperceptível à audição e à visão humana, mas não suportam atrasos. Inversamente, fluxos de dados de informações podem suportar pequenos atrasos, mas de maneira alguma podem conviver com perda nos dados.

As células são transmitidas através de conexões de circuitos virtuais, sendo seu encaminhamento baseado em informações do cabeçalho contido em cada uma delas, como será visto adiante.

Na hora da conexão, são estabelecidos diversos parâmetros de conexão. Como exemplo tem-se a taxa máxima de células (em horas de pico), se é uma conexão com taxa de transmissão variável ou fixa, etc. Também é negociado um parâmetro de Qualidade de Serviço (QoS), que determina o máximo de perda de células possível, o atraso máximo de transmissão, e assim por diante.

Essa tecnologia é utilizada tanto para formar um *backbone* de alta velocidade, como para suprir necessidades de redes locais de grande fluxo, ou seja, para ganho de desempenho no tráfego dos dados na rede. As velocidades de transmissão vão desde 25 ou 155 Mbps para redes locais, permitindo que se aproveite toda a estrutura já existente, como cabeamento e repetidores, chegando até 622 Mbps, para o que se exige o uso de fibras ópticas como meio de transmissão.

6.5.1. ATM versus STM

A principal diferença entre a transmissão síncrona (STM) e assíncrona (ATM) diz respeito à ocupação da banda. Observe a figura a seguir:

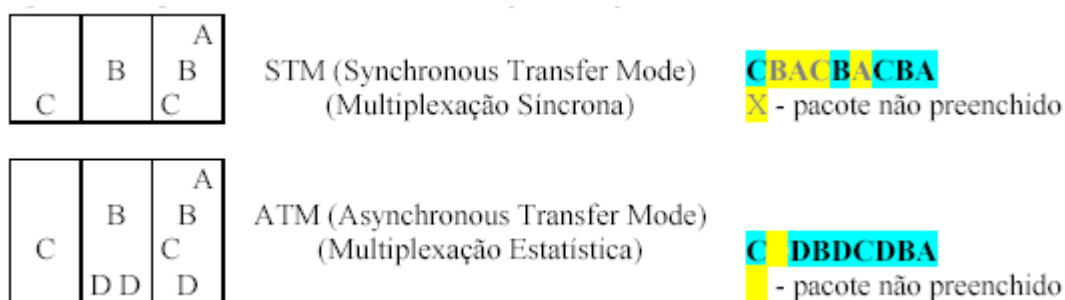



Figura 6-5.2: ATM vs. STM

No caso da multiplexação síncrona do STM, vê-se que ocorre uma má utilização da banda, pois mesmo que não haja informações a serem transmitidas, a janela de tempo no meio de transmissão é reservada, eliminando a possibilidade de outra estação utilizar o meio. Esse tipo de situação é bastante comum hoje em dia nas transmissões via concessionárias de

comunicação, pois praticamente todo o sistema brasileiro é baseado em STM, com as tecnologias PDH (*Plesiochronous Digital Hierarchy*) e SDH (*Synchronous Digital Hierarchy*).

Já no caso da multiplexação estatística do ATM, obtem-se uma melhor utilização da banda, pois a estação só transmite quando tem necessidade, ou seja, utiliza o meio físico sob demanda.

Adiante será visto com mais detalhe, mas vale a pena ressaltar que o ATM utiliza uma camada inferior síncrona (PDH, SDH, Células). Assim, ele consegue uma melhor utilização do meio físico com os mesmos recursos utilizados atualmente, somente baseando-se numa melhor forma de multiplexação (estatística). Isso quer dizer que o fluxo de células embaixo do ATM é constante, sendo elas utilizadas ou não.

6.5.2. A Camada Física

A camada física tem as seguintes funções: adaptação ou mapeamento do quadro de transmissão, dissociação das taxas de células, geração e verificação do HEC (*Header Error Control*), delimitação de células e a transmissão pelo meio físico.

A seguir, alguns desses ítens serão analisados com detalhes.

Adaptação ou mapeamento do quadro de transmissão

Para que a transmissão seja efetuada, deve haver um mapeamento do formato de quadro (SDH, PDH, Células) para a taxa de transmissão do meio físico, vista no item anterior. Assim, pode-se dizer que, sobre a taxa de transmissão do meio físico (STM1, STM4, E1, E3, Cell Based e assim por diante), é necessário um formato de quadro para organizar a informação. A figura a seguir mostra as alternativas possíveis de mapeamento no protocolo ATM.

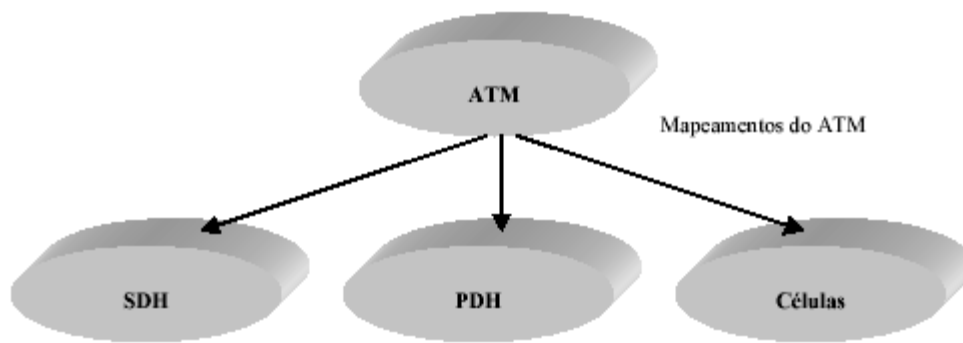



Figura 6-5.3: Mapeamento do quadro de transmissão

Pode-se demonstrar que, independentemente do mapeamento, a taxa de bits é sempre a mesma, ou seja, o overhead é constante nos vários mapeamentos. Exemplificando com o SDH, temos a seguinte figura:

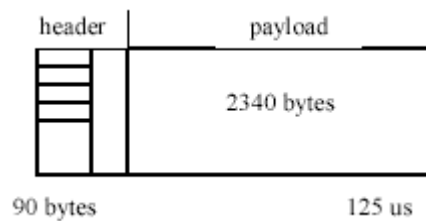



Figura 6-5.4: Exemplificando com o SDH

2340 + 90 = 2430 bytes

Taxa total de tx = 2430 bytes em 125 us = 19440 bits em 125us = 155,52 Mbps

Taxa de payload = 2340 bytes em 125 us = 18720 bits em 125 us = 149,76 Mbps

Além do overhead do STM, deve-se considerar o overhead da célula, que possui 5 bytes de header e 48 bytes de payload. Assim tem-se que a taxa útil de transmissão é:

Taxa útil de transmissão = 48/53 x 149,76 Mbps = 135,631698 Mbps

Células de informação / s = 135,631698 Mbps / (48 x 8) bits

= 353.208 células por segundo

Dissociação das taxas de células

Como já foi dito, temos uma transmissão assíncrona sobre um meio síncrono. Dessa forma, caso a camada de cima não tenha informações a transmitir, a taxa de células no meio físico deve permanecer constante. É função da camada física inserir na origem e retirar no destino células "idle" quando o ATM não fornece células úteis para transmissão. A figura a seguir ilustra o que foi dito.

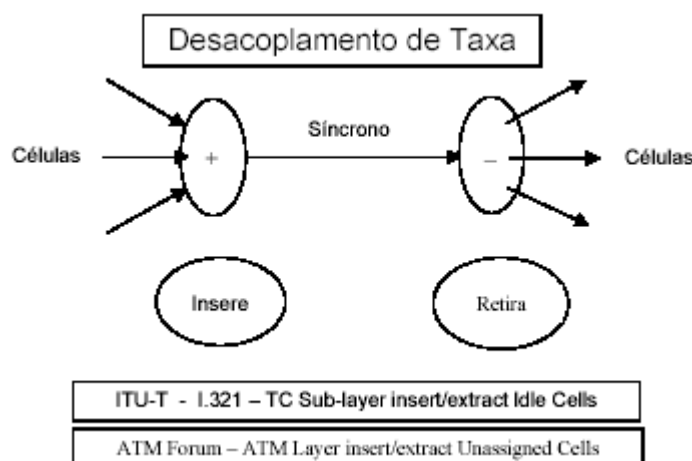



Figura 6-5.5: Desacoplamento de taxa

Geração e verificação do HEC (*Header Error Control*)

Caso tenha ocorrido erro em mais de um bit, descarta-se a célula. Ocorrendo erro em um único bit, corrige-se automaticamente e muda-se de estado. Se a próxima célula vier com erro, descarta-se. A figura a seguir ilustra o diagrama de estados.

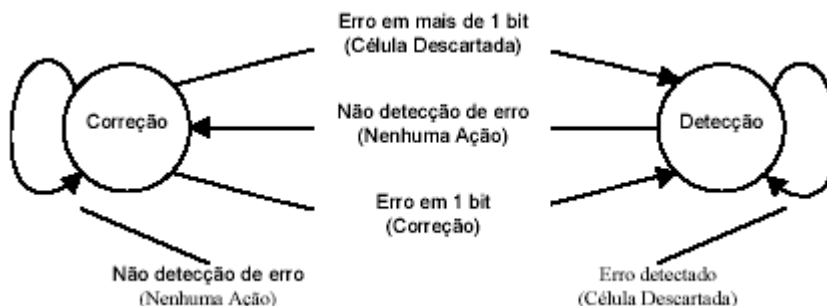



Figura 6-5.6: Diagrama de estados

Delineação de células

O sincronismo para delimitar o início da célula é feito através do HEC. Como o algoritmo é conhecido (CRC-8), e sabe-se que os quatro primeiros bytes do cabeçalho geram o último, basta ir lendo byte a byte até fechar. Então, testa-se célula a célula por determinado número de vezes. Se repetir, então sincronizou. A figura a seguir ilustra o que foi dito.

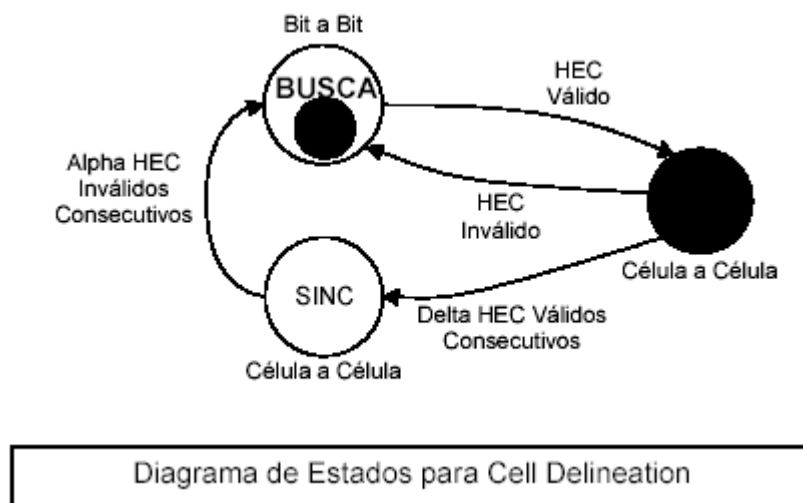



Figura 6-5.7: Diagrama de estados para Cell Delineation

6.5.3. A Camada ATM

As principais funções da camada ATM estão relacionadas ao cabeçalho da célula, e são:

- Construção da célula: adição do cabeçalho;
- Recepção da célula e validação do cabeçalho;
- Estabelecimento e liberação de conexões virtuais;
- Mux / Demux de conexões virtuais;
- Processamento na rede (roteamento, controle de tráfego e prioridade, suporte para sinalização, suporte para funções de OAM - fluxos f4 e f5).

Antes de analisar as funções da camada ATM, será necessário descrever a célula, seu cabeçalho e os diferentes tipos de células existentes.

6.5.4. A célula ATM

As células são de tamanho fixo de 53 *bytes*, com um cabeçalho que ocupa 5 *bytes*, e um campo de informação com 48 *bytes* de comprimento. Dessa forma, cada célula é identificada por seu cabeçalho, que contém informações indicando a conexão com o circuito virtual. A célula ATM pode ser representada como mostra a figura a seguir.

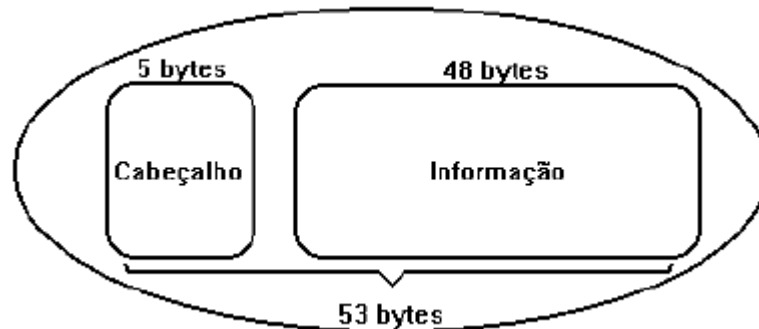



Figura 6-5.8: Célula ATM

Pelo fato das células serem de tamanho fixo, são diretamente responsáveis pelo enorme ganho de desempenho dos comutadores, terminais e dispositivos de comunicações. A grande razão para as células oferecerem esse ganho é porque elas podem ser processadas mais eficientemente que pacotes de tamanho variável ou "*bit-streams*". Além disso, através da segmentação do tráfego em tamanho fixo, as células se tornam um poderoso mecanismo que pode multiplexar tráfegos de diferentes características sobre uma estrutura comum de comunicações.

O Retardo de Empacotamento

Para exemplificar o efeito do tamanho da célula no retardo de transmissão, tem-se que voz PCM padrão, a 64Kbps envia 1 byte a cada 125us. Assim, para fechar 48 bytes gasta-se 6ms, que é o retardo de empacotamento. A mesma coisa acontece na chegada da célula no destino, ou seja, voz já atrasa 12 ms devido ao retardo de empacotamento, fora o retardo de propagação.

Maior payload: maior retardo de empacotamento - ruim para voz e bom para dados

Menor payload: menor retardo de empacotamento - bom para voz e ruim para dados

Compromisso voz / dados

Com base no que foi visto acima, formaram-se dois grupos, o de voz, que queria um payload menor, de 32 bytes, e o grupo de dados, que queria um payload maior, de 64 bytes. Dizem que foi tomada uma decisão salomônica, e o payload atual de 48 bytes é a média aritmética entre os dois ((64+32)/2).

6.5.5. O cabeçalho UNI (User Network Interface)

O ATM é orientado à conexão e o meio é confiável, portanto, as funcionalidades do cabeçalho ATM são reduzidas:

- Não identifica fonte e destino (utiliza identificadores nó a nó);
- Não identifica a seqüência de células para efeito de remontagem (se sair de seqüência, dá erro nas camadas superiores - chega tudo trocado);
- Controle de erros e retransmissão somente no destino;
- No cabeçalho, somente a geração do HEC é função da camada física, o resto é função da camada ATM. Nos 5 *bytes* de cabeçalho da célula, encontram-se os campos para comutação (através do par VPI e VCI), identificação do tipo, prioridade para descarte e um verificador de erros do cabeçalho, como mostra a figura a seguir. No caso da célula ser utilizada em uma interface rede-rede, não há o campo para o controle de fluxo genérico (GFC), aumentando o tamanho destinado ao VPI.

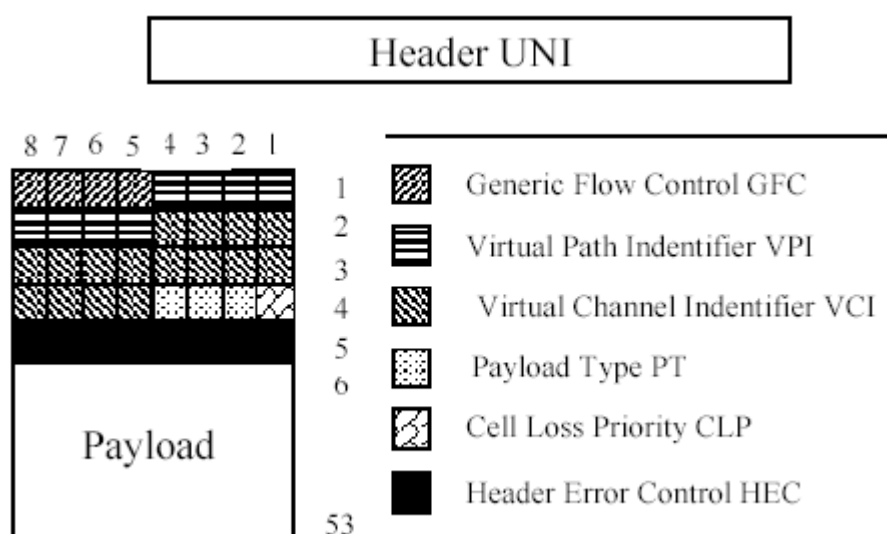



Figura 6-5.9: Header UNI

Na transmissão, o primeiro byte a ser enviado é o número 1, bit 8 (no caso NNI, VPI bit 8), a segunda é o VPI bit 7, e assim por diante, da direita para a esquerda e de cima para baixo.

O significado dos campos das células está descrito nos itens a seguir.

GFC (*Generic Flow Control*)

Quando introduzido, serve para regular o fluxo em uma rede ATM, ou seja, evita condições de sobrecarga nas interfaces do usuário, mas não realiza controle de fluxo sobre o tráfego vindo da rede. Entretanto, sua utilização ainda não foi padronizada, existindo algumas alternativas possíveis, que podem ser encontradas em [25].

VPI (*Virtual Path Identifier*) e *VCI (Virtual Channel Identifier*)

As transmissões numa rede ATM são efetuadas através de conexões. Uma conexão fim a fim é chamada **conexão com canal virtual** (*Virtual Channel Connection* - VCC). Cada conexão virtual em um enlace é denominada de **enlace de canal virtual** (*Virtual Channel link* – VCL). A figura a seguir ilustra isso.

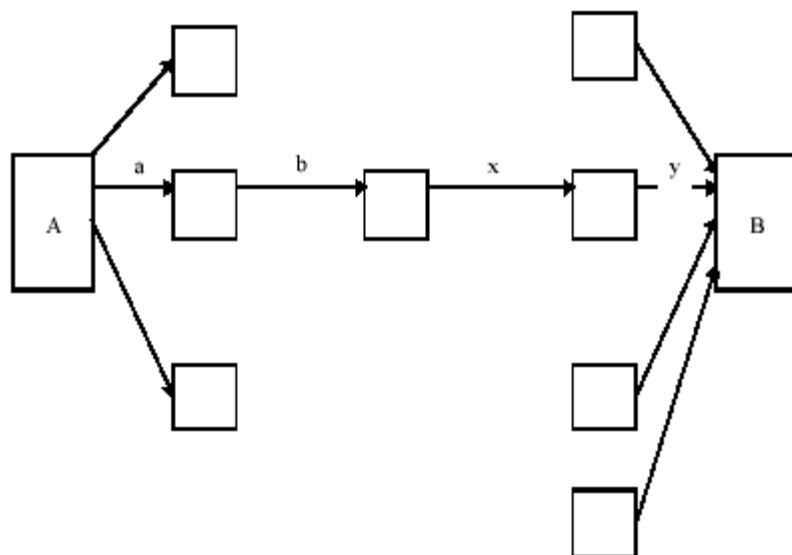



Figura 6-5.10: Enlace de canal virtual

Existe uma VCC entre os pontos A e B da figura acima, formada por quatro VCLs, identificadas pelos rótulos a, b, x e y. Quando A e B estabeleceram a conexão, as tabelas de cada nó intermediário foram atualizadas para redirecionarem as células corretamente.

Quando uma célula chega a um comutador, este identifica o caminho que está registrado no cabeçalho da célula (par VPI e VCI), e consulta um tabela de acordo com a porta de entrada dessa célula, para redirecioná-la a uma porta de saída. Antes da retransmissão, porém, é necessário atualizar o cabeçalho da célula de acordo com o próximo enlace de canal.

A figura a seguir ilustra a comutação através de rótulo [24]. Na figura, pode-se ver um único rótulo redirecionando a célula, mas é bom lembrar que existem dois rótulos em cada célula, que serão melhor explicados adiante.

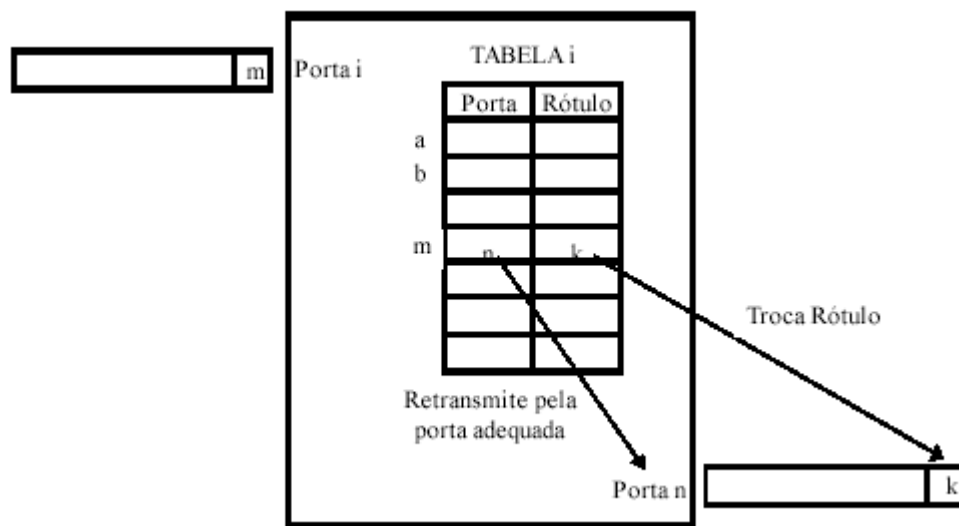



Figura 6-5.11: Comutação através de rótulo

Quando a célula enviada pela porta n (com o rótulo k) chega no outro comutador, o processo é semelhante, ou seja, a tabela da porta por onde chegar a célula será analisada no índice k, retransmitindo a célula pela porta adequada. Esse processo se repete até chegar ao destino final.

PT (*Payload Type*)

Contém o tipo do conteúdo que consta no campo de informação da célula, indicando se são dados do usuário ou informações de gerência. No item "Tipo de Células" adiante, esse campo será visto com detalhes.

CLP (*Cell Lost Priority*)

Indica a prioridade de descarte das células em caso de congestionamento da rede. Se este bit estiver ligado, a célula será escolhida para descarte caso necessário.

HEC (*Header Error Control*)

CRC X8+X2+X+1. Contém uma seqüência de bits obtida a partir das informações do cabeçalho, de modo a permitir que o receptor verifique a integridade do mesmo. Esse campo é preenchido pela subcamada de convergência da camada física, e não é responsabilidade da camada ATM.

É importante observar que não existe qualquer campo na célula que garanta a integridade dos dados. Contudo, é de responsabilidade das camadas superiores a verificação e eventual solicitação de retransmissão de alguma informação que tenha sido corrompida. O objetivo é simplificar o processamento em cada nó intermediário, amparado no fato de se ter uma grande confiabilidade no meio de transmissão, que normalmente é a fibra óptica.

6.5.6. A Camada AAL

A camada de adaptação (AAL - ATM *Adaptation Layer*) situa-se antes da camada ATM e é responsável pela segmentação dos serviços de seus formatos nativos (originais), para células ATM de tamanho fixo. Diferentes tipos de adaptações são necessárias para os diferentes tipos de serviços prestados.

A camada de adaptação recebe a informação do tipo de serviço a ser transmitido pela rede, alguns de taxa variável, outros de taxa fixa de transmissão, e segmenta essas informações nas células. Essa informação do tipo de serviço que deve ser oferecido vem desde onde foi originado o serviço até a camada de adaptação, que o tratará conforme a necessidade da classe em que ele se enquadra. No recebimento, extrai as informações das células e as remonta em seu formato original.

O AAL não é caracterizado por um conjunto bem definido de funções, ele deve suportar quaisquer funções que forem solicitadas por qualquer protocolo que utilize seus serviços.

Podemos exemplificar como serviços por ele oferecidos:

- Recuperação de erros de transmissão;
- Tratamento de perdas e duplicações de células;
- Controle de fluxo e controle de sincronismo;
- Efetua a quebra e remontagem de quadros em células.

Classes de Serviços

Os serviços foram divididos em classes de tráfego, como pode ser visualizado na tabela 6-5.1, seguindo parâmetros como a sincronização entre a origem e o destino (se é ou não necessária), a taxa de transmissão (se é variável ou fixa) e se é ou não orientado à conexão.

Parâmetro	Classe A	Classe B	Classe C	Classe D
Sincronização entre origem e destino	Necessária		Não necessária	
Taxa de transmissão	Constante	Variável		
Modo de conexão	Orientado à conexões			Não orientado a conexões
Exemplo de serviços	Emulação de circuitos, vídeo a taxa constante	Vídeo e áudio a taxas variáveis (comprimidos)	Transferência de dados orientada à conexões	Transferência de dados não orientada à conexões
Protocolo AAL	AAL 1	AAL 2	AAL 3, 4	AAL 5

Tabela 6-5.1: Classificação dos serviços

Alguns requisitos dos serviços de classes A e B, que têm como exemplo vídeo a taxa constante (A) e vídeo e áudio a taxa variável (B), são:

- Pequeno retardo máximo de transferência (no pior caso, idêntico ao atraso que haveria se um cabo ligasse diretamente os dois pontos);
- Variação de atraso desprezível;

- Transporte de dois bits/bytes com manutenção do intervalo entre eles;
- Tratamento adequado de perdas, duplicações e erros em células recebidas;
- Tratamento do relógio para manter o correto tempo entre as células.

Alguns requisitos dos serviços de classes C e D, que têm como exemplo transferência de dados orientada a conexões (C) e transferência de dados não orientada à conexões (D), são:

- Retardo máximo de transferência moderado;
- Variação moderada do atraso;
- Ausência de requisito de sincronização entre unidades de dados.

6.6. SDH

Rede SDH é o conjunto de equipamentos e meios físicos de transmissão que compõem um sistema digital síncrono de transporte de informações. Este sistema tem o objetivo de fornecer uma infra-estrutura básica para redes de dados e voz, e atualmente é utilizado em muitas empresas que prestam serviços de Telecomunicações, públicas e privadas, em todo o mundo.

As tecnologias SDH (Synchronous Digital Hierarchy) são utilizadas para multiplexação TDM com altas taxas de bits, tendo a fibra óptica como meio físico preferencial de transmissão. Entretanto, possui ainda interfaces elétricas que permitem o uso de outros meios físicos de transmissão, tais como enlaces de rádios digitais e sistemas ópticos de visada direta, que utilizam feixes de luz infravermelha.

Sua elevada flexibilidade para transportar diferentes tipos de hierarquias digitais permite oferecer interfaces compatíveis com o padrão PDH europeu (nas taxas de 2 Mbit/s, 8 Mbit/s, 34 Mbit/s e 140 Mbit/s) e americano (nas taxas de 1,5 Mbit/s, 6 Mbit/s e 45 Mbit/s), além do próprio SDH (nas taxas de 155 Mbit/s, 622 Mbit/s, 2,5 Gbit/s e 10 Gbit/s).

A tecnologia SDH permite ainda implementar mecanismos variados de proteção nos equipamentos e na própria rede, oferecendo serviços com alta disponibilidade e efetiva segurança no transporte de informações.

6.6.1. Histórico

Os primeiros sistemas de transmissão baseados em fibra óptica utilizados nas redes de telefonia pública utilizavam tecnologias proprietárias na sua arquitetura, nos formatos de multiplexação, no software e no hardware, e tinham procedimentos de manutenção diferenciados. Os usuários desses equipamentos solicitaram ao mercado fornecedor que desenvolvesse uma padronização de tecnologias e equipamentos de forma a possibilitar a utilização de equipamentos de diferentes fornecedores numa mesma rede.

A tarefa de criar tais padrões começou em 1984, junto com outras frentes de trabalho para outras tecnologias, e ficou inicialmente a cargo da ECSA - EUA (*Exchange Carriers Standards Association*). A ECSA desenvolveu o padrão SONET (*Synchronous Optical Network*), que foi adotado, entre outros países, pelos EUA.

Após algum tempo o ITU-T - Europa (antigo CCITT) envolveu-se no trabalho para que um único padrão internacional pudesse ser desenvolvido a fim de criar um sistema que possibilitasse que as redes de telefonia de distintos países pudessem ser interligadas. O resultado desse trabalho foi o conjunto de padrões e recomendações conhecido como SDH (Synchronous Digital Hierachy), ou Hierarquia Digital Síncrona.

O desenvolvimento do SDH levou a um ajuste no padrão SONET para que os frames do 2 sistemas pudessem ser compatíveis tanto em tamanho como em taxa de bits, de forma que se pudessem interligar a redes dos 2 padrões sem problemas de interface.

A tabela a seguir mostra a relação entre os sinais SONET e SDH.

SONET	Taxa de Bits (kbit/s)	SDH
STS1, OC1	51 840	STM0
STS3, OC3	155 520	STM1
STS12, OC12	622 080	STM4
STS48, OC48	2 488 320	STM16
STS192, OC192	9 953 280	STM64

Tabela 6-6.1: Tabela relacionando SONET e SDH

6.6.2. Rede SDH

Uma rede SDH é composta por:

- Rede Física: é o meio de transmissão que interliga os equipamentos SDH. Pode ser composta por: cabos de fibra óptica, enlaces de rádio e sistemas ópticos de visada direta baseados em feixes de luz infravermelha.
- Equipamentos: são os multiplexadores SDH de diversas capacidades que executam o transporte de informações.
- Sistema de Gerência: é o sistema responsável pelo gerenciamento da rede SDH, contendo as funcionalidades de supervisão e controle da rede, e de configuração de equipamentos e provisionamento de facilidades.
- Sistema de Sincronismo: é o sistema responsável pelo fornecimento das referências de relógio para os equipamentos da rede SDH, e que garante a propagação desse sinal por toda a rede.

A figura a seguir apresenta um exemplo de rede SDH.

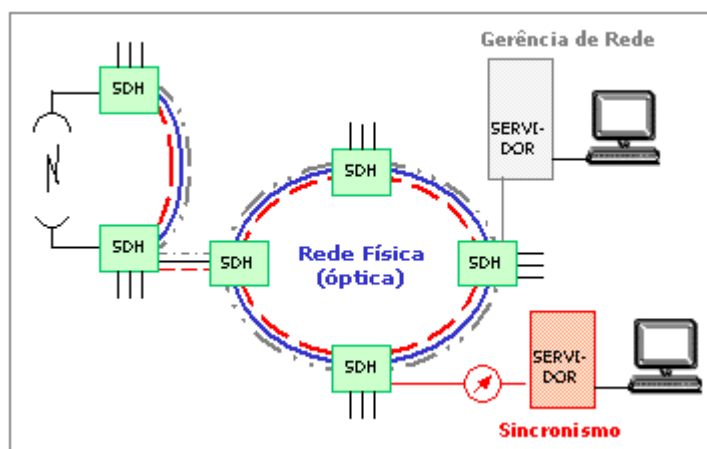



Figura 6-6.1: Exemplo de rede SDH

Vantagens e restrições

As redes SDH oferecem vários benefícios, quando comparadas com outras tecnologias:

- O cabeçalho complexo existente no frame SDH permite a gerência (administração, operação e manutenção) centralizada da rede;
- A arquitetura de multiplexação síncrona e a padronização tanto em nível de equipamentos como de interfaces, permite o crescimento para níveis mais altos de multiplexação e taxas de bits;
- A estrutura de multiplexação é flexível, permitindo o transporte de sinais PDH (e até mesmo de células ATM) e o acesso aos tributários de qualquer hierarquia num único equipamento;
- A forte padronização do SDH permite maior compatibilidade entre equipamentos de fabricantes diferentes tanto através de interfaces elétricas como ópticas;
- Os equipamentos possuem mecanismos que permitem implementar procedimentos de proteção tanto nas interfaces de tributários como na rede, facilitando a formação de redes em anel ou malha.

Entretanto, a tecnologia SDH apresenta ainda as seguintes desvantagens:

- O projeto, instalação e operação da rede SDH é complexo e deve ser feito com um planejamento criterioso e detalhado;

- Apesar da forte padronização de equipamentos e da tecnologia SDH, a padronização dos sistemas de gerência de rede ainda não é um fato, impedindo que equipamentos de fabricantes diferentes possam ser gerenciados por um sistema único.

6.6.3. Características do SDH

A hierarquia SDH foi concebida para uma arquitetura de multiplexação síncrona. Cada canal opera com um relógio sincronizado com os relógios dos outros canais, e é sincronizado com o equipamento multiplex através de um processo de justificação de bit e encapsulamento da informação (contêiner). A esse contêiner é adicionado um cabeçalho (POH), que o caracteriza e indica sua localização no frame, e forma-se então um contêiner virtual (VC - Virtual Container) para cada canal.

O SDH pode transportar também os diferentes tipos de sinais PDH, através do frame padronizado denominado STM-N (Syncronous Transport Module), utilizado tanto para sinais elétricos como para sinais ópticos. Atualmente o padrão SDH utiliza frames STM-N com as seguintes taxas de bits: 155520 Mbit/s (STM-1 elétrico ou óptico), 622080 Mbit/s (STM-4 óptico), 2488320 Mbit/s ou 2,5 Gbit/s (STM-16 óptico) e 9953280 Mbit/s ou 10 Gbit/s (STM-64 óptico).

Os diversos canais multiplexados (VC's) normalmente são chamados de tributários, e os sinais de transporte gerados (STM-N) são chamados de agregados ou sinais de linha.

Os sub-ítens a seguir detalham as características mais relevantes da tecnologia SDH:

Sincronismo

As redes SDH formam um sistema síncrono onde todos os relógios de seus equipamentos têm, em média, a mesma freqüência. O relógio de cada equipamento, chamado de relógio secundário ou escravo, pode ser rastreado até o relógio principal da rede, chamado também de mestre, garantindo a distribuição e qualidade do sinal de sincronismo.

A manutenção de uma boa referência de relógio permite que os sinais STM-1 mantenham sua taxa de 155 Mbit/s estável, e que vários sinais STM-1 síncronos possam ser multiplexados sem a necessidade de inserção de bits, sendo facilmente acessados em sinais STM-N de maior taxa de bits.

Também os sinais síncronos de menores taxas de bits, encapsulados nos VC's, podem ser multiplexados sem a necessidade de inserção de bits para compor os sinais STM-1, e podem ser facilmente acessados e recuperados.

O uso de ponteiros em conjunto com buffers permite acomodar as eventuais diferenças de fase e freqüência dos canais durante o processo de multiplexação. Os ponteiros possuem campos específicos para armazenar os bits ou bytes em excesso ou para indicar a falta destes durante o processo de sincronização (justificação). Os buffers permitem que esse processo ocorra sem a perda de informação armazenando e mantendo o sinal original.

Desta forma, é extremamente importante a qualidade e a manutenção do sinal de sincronismo para o sucesso da rede e dos serviços prestados a partir dela.

Estrutura em Camadas

O padrão SDH foi desenvolvido usando a abordagem cliente/servidor e sua arquitetura de administração e supervisão procurou apoiar-se no modelo de camadas OSI (ISO), permitindo que a supervisão do transporte de informações seja feita através de camadas hierarquizadas.

Do ponto de vista de rede, essas camadas são representadas conforme a figura a seguir. Para um determinado serviço, caracterizado por sua origem e destino e por uma taxa de bits conhecida, são identificados os tipos de funcionalidades e as camadas envolvidas para executa-lo.

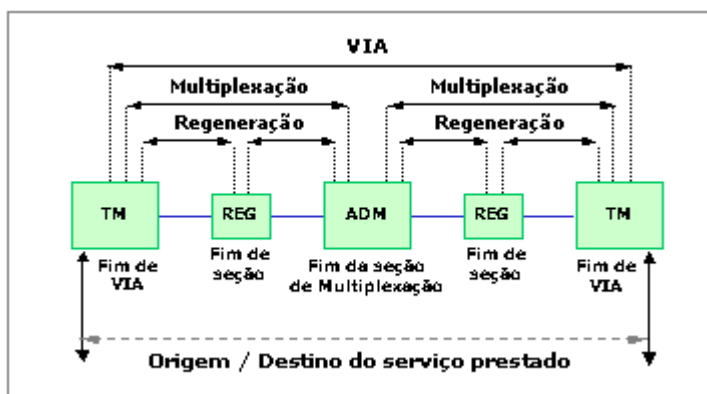



Figura 6-6.2: Rede em camadas

Entende-se por Via o caminho percorrido pelo sinal entre a origem e o destino. Nesse caminho o sinal é acondicionado no frame SDH que faz o seu transporte através de todos

os equipamentos da rede nessa rota. Em cada equipamento, de acordo com a sua função, o frame é processado pelas camadas adequadas para ser restaurado ou para extrair ou inserir novos serviços. Em cada etapa desse processo a informações de administração e supervisão do SDH são geradas e inseridas no frame.

O modelo em camadas para um determinado equipamento da rede é apresentado na figura a seguir.

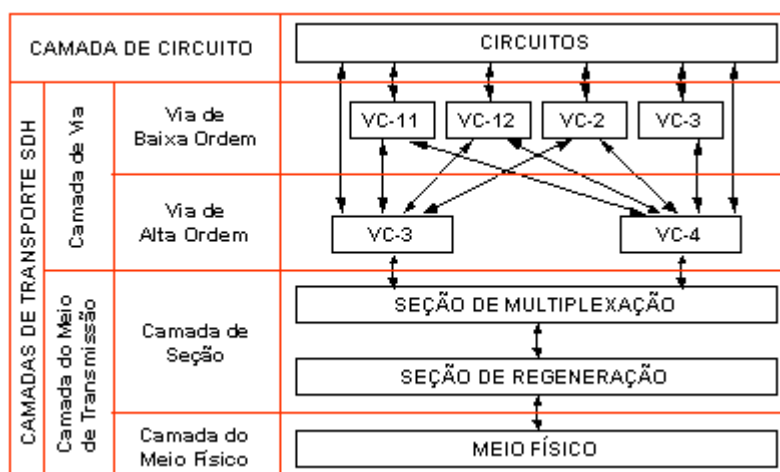



Figura 6-6.3: Modelo de camadas

A camada do meio de transmissão é dependente do meio utilizado, e por isso foi dividida em 2 camadas distintas: meio físico e seção. A camada do meio físico realiza o condicionamento do sinal de acordo com esse meio, seja ele óptico ou elétrico.

A camada de seção também está dividida em 2 novas camadas. A sub-camada de regeneração é responsável pelo processamento dos frames em todos os equipamentos da rede, sejam eles de passagem, de extração ou inserção de tributários, ou de terminação de via. A seção de multiplexação é responsável pelo processamento fim-a-fim dos frames nos equipamentos de extração ou inserção de tributários, ou de terminação de via.

A camada de via está divida em alta ordem e baixa ordem. Nessa camada cada VC é uma estrutura com a informação útil (contêiner) e um cabeçalho que o caracteriza (POH). Na via de baixa ordem cada VC contém um único contêiner e seu cabeçalho (VC-1x, VC-2 ou VC-3). Na via de alta ordem um VC pode conter um único contêiner e seu cabeçalho (VC-3 ou VC-4), ou um conjunto de contêineres de menor ordem e o respectivo cabeçalho.

A camada de circuito realiza o condicionamento da informação útil retirada do contêiner para a interface elétrica ou óptica definida para cada serviço a ser fornecido pelo equipamento.

Estrutura do Frame

O frame SDH tem tamanho padrão para cada hierarquia. Cada frame constitui uma unidade para fins de administração e supervisão da transmissão no sistema. Esses frames são transmitidos a uma taxa de 8000 frames por segundo (8000 Hz).

O frame SDH para a hierarquia STM-1, por exemplo, tem 2430 bytes, organizados em 9 linhas com 270 colunas de bytes, os quais são transmitidos serialmente linha a linha da esquerda para a direita, e de cima para baixo. Sua estrutura básica é apresentada na figura a seguir.

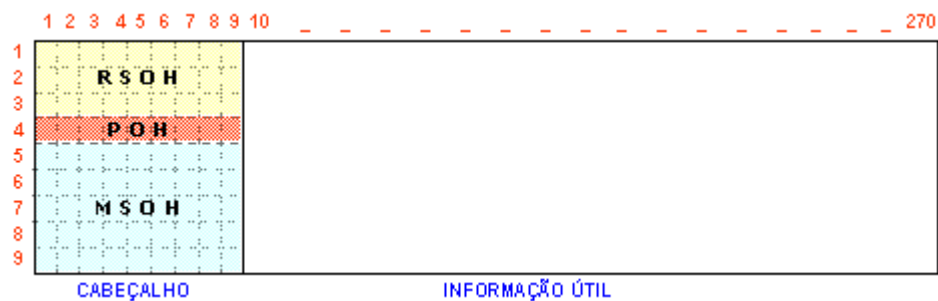



Figura 6-6.4: Estrutura do Frame

O cabeçalho (overhead) é composto por 3 tipos de estruturas:

- RSOH (Regenerator Section Overhead), processado em cada equipamento da rede, contém informações de alinhamento de frame, identificação de frame, monitoração de erro de regeneração, alarmes físicos externos ao equipamento, e supervisão de sistema. Contém também um canal de voz, para comunicação de técnicos entre equipamentos.
- MSOH (Multiplex Section Overhead), processado apenas em equipamentos onde existe inserção (add) ou retirada (drop) de canais multiplexados, contém informações de monitoração e indicação de erros de multiplexação, controle de chaveamento de mecanismos de proteção, monitoração de sincronismo e gerência de sistema.
- POH (Path Overhead), processado em cada equipamento, possui os ponteiros que indicam onde se localiza o primeiro byte do(s) VC(s) dentro da área de informação útil (payload) do frame, e eventuais bytes provenientes de justificação desse(s) VC(s).

A incorporação dos ponteiros nas estruturas dos VC's do frame SDH permite que mesmos sinais com diferenças de fase e freqüência possam ser transportados num mesmo frame, já que essas diferenças são acomodadas em bytes específicos do POH através do processo de justificação. Ressalta-se, entretanto, que essas diferenças devem atender às especificações estabelecidas pelas recomendações do ITU-T para o SDH.

Processo de Multiplexação

A figura a seguir apresenta o processo de multiplexação dos canais tributários no frame SDH.

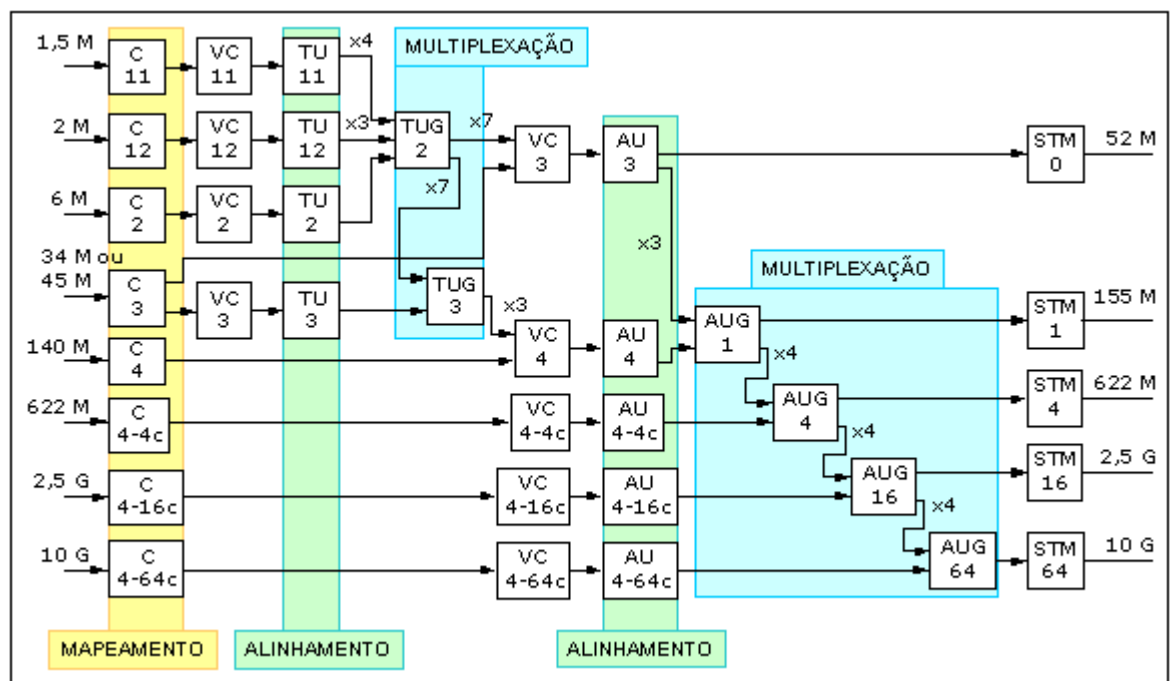



Figura 6-6.5: Processo de multiplexação dos canais do frame SDH

O processo de multiplexação dos canais tributários no frame SDH tem os seguintes passos:

- Mapeamento, onde os tributários são sincronizados com o equipamento multiplex (justificação de bit), encapsulados e recebem seus ponteiros (POH) para formar os VC's;
- Alinhamento, onde os VC's recebem novos ponteiros para formarem as unidades TU (Tributary Unit) ou AU (Administrative Unit), para permitir que o primeiro byte do VC seja localizado;

- Multiplexação byte a byte, onde os VC's de baixa ordem são agrupados para compor os VC's de alta ordem ou os VC's de alta ordem são processados para formar os AUG (Administrative Unit Group);
- Preenchimento, onde, na falta de tributários configurados ou para completar o espaço restante de tributários de baixa ordem, são adicionados bits sem informação para completar o frame.

Nos equipamentos do padrão SDH o processo de multiplexação normalmente é executado pela matriz de conexão cruzada (Cross-connect Matrix). A capacidade desta matriz para compor os frames SDH com canais de taxas de bits diversas define, de fato, a capacidade do equipamento.

Normalmente os equipamentos com sinais agregados de taxas de bits até STM-4 (622 Mbit/s) possuem matrizes com capacidade para multiplexar canais com taxa de bits de 2 Mbit/s até 155 Mbit/s. Os equipamentos com sinais agregados de taxas de bits superiores a STM-4 (622 Mbit/s) possuem matrizes com capacidade para multiplexar canais com taxa mínima de 155 Mbit/s.

6.6.4. Equipamentos

O padrão SDH definiu 3 tipos de equipamentos para compor a rede:

- TM (Terminal Multiplex): possui apenas uma interface de agregados e possibilita a inserção (add) ou retirada (drop) de tributários de diversas hierarquias;
- ADM (Add and Drop Multiplex): possui duas interfaces de agregados e possibilita a inserção (add) ou retirada (drop) de tributários de diversas hierarquias. Estes equipamentos também podem ser usados como regeneradores de sinal, quando nenhuma interface de tributário é instalada.
- SDXC (Synchronous Digital Cross-connect): possui interfaces de entrada e saída de diversas hierarquias e pode interligá-las com uma grande infinidade de combinações.

A figura a seguir apresenta esses equipamentos.

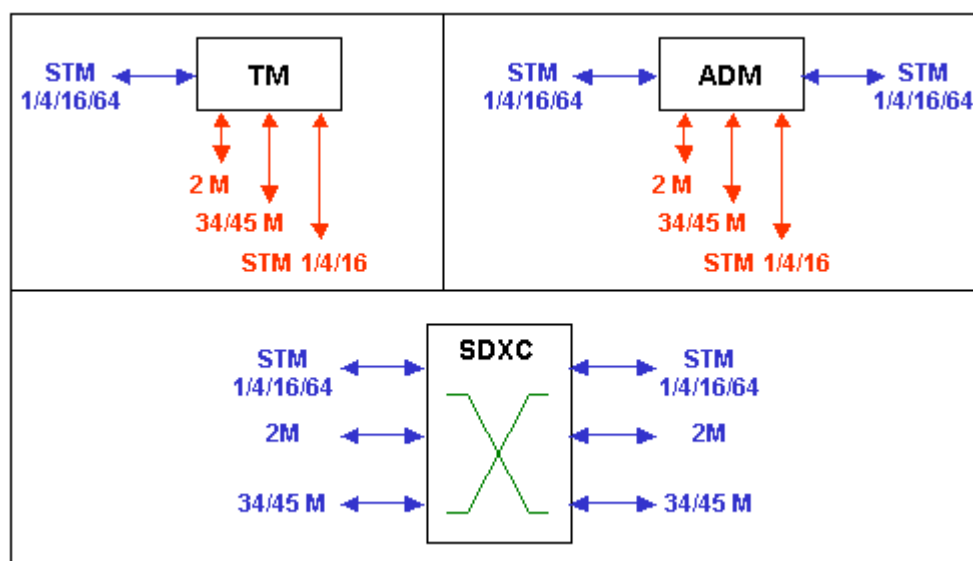



Figura 6-6.6: Equipamentos

Embora esses tipos de equipamentos tenham sido especificados nas recomendações do ITU-T, com detalhes de blocos funcionais, os fabricantes de equipamentos fornecem, em sua maioria, apenas os ADM's, que podem executar a função de ADM e de TM com diversas capacidades de taxas de bits, e os SDXC, também com diversas possibilidades de configuração.

Para selecionar e utilizar esses equipamentos em redes SDH devem ser considerados os seguintes aspectos:

- Tributários: tipos (elétricos, ópticos), taxas de bits, número de interfaces por placas e número máximo de placas no equipamento;
- Agregados: tipos (elétricos, ópticos), taxas de bits e número máximo de placas no equipamento;
- Matriz de Conexão Cruzada (Cross-connect Matrix): capacidade total da matriz e taxas de bits do canais a serem multiplexados.

6.7. MPLS (*Multi-protocol label switching)*

Nos últimos anos, com a crescente popularização da Internet, foi observada uma expansão inédita das redes. As grandes corporações, bem como pessoas físicas, passaram a depender, de forma muito mais intensa, de serviços de rede. O aparecimento de necessidades de confiança, eficiência e qualidade de serviço, em relação às redes utilizadas acompanhou este crescimento. Os provedores de serviços de Internet estão, mais do que nunca, analisando de forma crítica qualquer aspecto relacionado ao ambiente operacional, procurando por oportunidades de crescimento e otimização de performance. Além disso, a crise que vem afetando o setor no último ano, obriga as companhias provedoras de acesso a procurarem novos serviços para incrementar suas receitas.

Nesse contexto, a engenharia de tráfego emerge como uma consideração de maior importância no desenho e operação de grandes redes públicas de Internet. Entretanto, os protocolos de roteamento clássicos da Internet impedem a realização de políticas avançadas de engenharia de tráfego em redes IP legadas. O advento do MPLS (*Multi-protocol label switching)* acena com a possibilidade de atender a algumas dessas necessidades.

Neste item, apresenta-se a tecnologia em que se baseia o MPLS, incluindo suas capacidades e limitações. Desta forma, pretende-se munir o leitor de informações necessárias à formulação de conclusões racionais sobre as possíveis aplicações de MPLS, bem como suas limitações e futuro mercadológico.

6.7.1. Surgimento, Evolução e Áreas de aplicação

O MPLS foi inicialmente concebido para acelerar a expedição de pacotes em roteadores legados baseados em software, mas acabou resultando em importantes avanços na tecnologia de plano de controle IP, engenharia de tráfego, redes privadas virtuais (VPNs) e gerenciamento de conexões em redes óticas. Quando o MPLS é combinado com serviços diferenciados e roteamento baseado em restrições, diversos tipos de Qualidade de Serviço (QoS) podem ser implementados em redes IP.

O principal conceito de arquitetura que suporta o MPLS é a separação clara do *plano de controle* e do *plano de dados*. O *plano de dados* consiste nos componentes de reenvio que promovem uma simples comutação por rótulos. O *plano de controle* está preocupado com

funções de coordenação, como roteamento e sinalização, para facilitar o deslocamento do tráfego através de toda a rede.

Uma das primeiras aplicações de MPLS em redes IP operacionais foi na engenharia de tráfego. Neste caso, é enfatizada a otimização da rede, com objetivos relacionados a QoS como menor atraso, menor variação de atraso, alta taxa de transmissão, pequena perda de pacotes e serviço previsível. Por outro lado, a otimização também se preocupa com minimizar os custos aos provedores pela utilização eficiente de recursos da rede. A aplicação de MPLS nesse campo veio das limitações impostas pelos protocolos utilizados, baseados em algoritmos de roteamento de menor caminho. O maior problema desses protocolos é não levar em conta limitações de capacidade ou natureza do tráfego. O resultado é o congestionamento de alguns segmentos da rede, enquanto permanecem subutilizados.

Outra importante aplicação do MPLS, atualmente em consideração, diz respeito ao gerenciamento de QoS em redes IP. O MPLS por si só não provê QoS, entretanto, quando combinado com roteamento baseado em restrições e serviços diferenciados, permite sofisticadas capacidades nesse sentido.

Uma outra importante aplicação é relacionada a VPNs. Tipicamente, VPNs aplicam técnicas de *tunelamento* para isolar o tráfego pertencente à rede privada do resto do tráfego na rede. Para esta aplicação, o MPLS pode ser visto como uma tecnologia de tunelamento que suporta a implementação de serviços de VPN.

Por último, o plano de controle foi estendido e generalizado para servir como plano de controle para tipos diferentes de redes de transporte comutadas, desde redes comutadas a pacotes e tecnologias baseadas em *multiplexação por divisão no tempo* (TDM), a redes de transporte óticas comutadas automaticamente. Esse plano de controle genérico está sendo padronizado pelo IETF dentro do conceito de MPLS generalizado (GMPLS).

6.7.2. Conceitos e funcionamento

Muitos conceitos básicos aplicados a qualquer tecnologia de encaminhamento devem ser revisados antes de descrever o funcionamento do MPLS.

Roteamento (*routing)*:

É o termo utilizado para descrever as ações tomadas por qualquer rede para transmitir pacotes. Pode haver um número qualquer de roteadores em uma rede, conectados de forma arbitrária. Os pacotes avançam na rede sendo enviados de uma máquina à outra em direção ao seu destino. Protocolos de roteamento permitem a cada máquina entender qual é deve ser o próximo "nó" em que um pacote deve passar para chegar ao seu destino. Os roteadores utilizam protocolos de roteamento para construir as chamadas "tabelas de roteamento". Quando um pacote é recebido e deve ser tomada uma decisão de encaminhamento, o roteador procura nesta tabela, utilizando o endereço de destino do pacote como um índice, obtendo qual deve ser a próxima máquina. A construção das tabelas e seu uso são operações lógicas essencialmente independentes. A figura abaixo ilustra essas funções como podem ocorrer em um roteador.

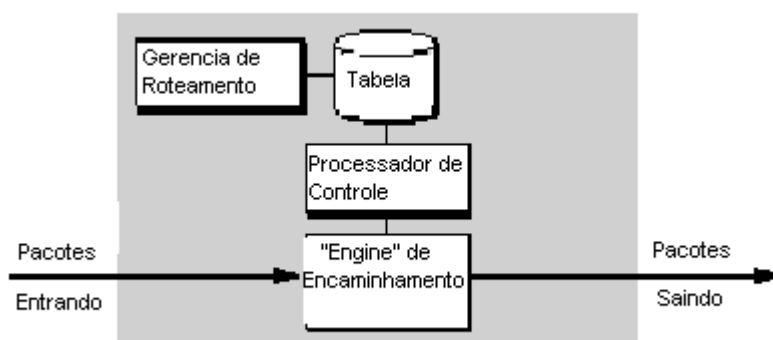



Figura 6-7.1: Funções do roteador

Comutação (*switching):*

É geralmente usada para descrever a transferência de dados de uma porta de entrada para uma porta de saída de uma máquina onde a seleção é baseada em informações de camada 2 (por exemplo, ATM VPI/VCI).

Componente de controle:

Constrói e mantém uma tabela de encaminhamento para uso de um nó. Trabalha com os componentes de controle de outros nós para distribuir informações de roteamento de forma consistente e acurada. Também certifica que procedimentos locais consistentes são usados para criar as tabelas de encaminhamento. Protocolos de rotamento padrão (OSPF, BGP, RIP) são usados para trocar informações de roteamento entre os componentes de controle. Deve reagir quando alguma alteração ocorre na rede, como falha de um enlace, mas não está envolvido no processamento de pacotes individuais.

Componente de encaminhamento:

Realiza o encaminhamento de pacotes propriamente dito. Usa informações da tabela de encaminhamento, informações contidas no próprio pacote e um conjunto de procedimentos locais para tomar a decisão. Em um roteador convencional, um algoritmo de busca compara o endereço de destino com entradas na tabela de encaminhamento até obter a informação mais próxima possível. É importante ressaltar que um processo de tomada de decisão deve ser repetido em cada nó ao longo do caminho da origem até o destino. Em um roteador de tecnologia MPLS (LSR), um algoritmo de busca exata e troca de rótulo usa o rótulo no pacote e uma tabela de encaminhamento baseada em rótulos para obter um novo rótulo e uma interface de saída para o pacote.

Tabela de encaminhamento:

É um conjunto de registros em uma tabela que provê informação para ajudar o componente de encaminhamento a executar sua função. A tabela de encaminhamento deve associar cada pacote com um registro que provê instruções relativas ao próximo destino do pacote.

Classe de equivalência de encaminhamento (FEC):

É definida como qualquer grupo de pacotes que podem ser tratados de forma equivalente para fins de encaminhamento. Exemplo de uma FEC é um conjunto de pacotes vindos de uma mesma origem com endereços de destino que tenham o mesmo prefixo de endereço IP. Outra FEC é o conjunto de pacotes cujos endereços de origem e destino são os

mesmos. FECs podem ser definidas para qualquer nível de granularidade. A figura abaixo demonstra esta idéia:

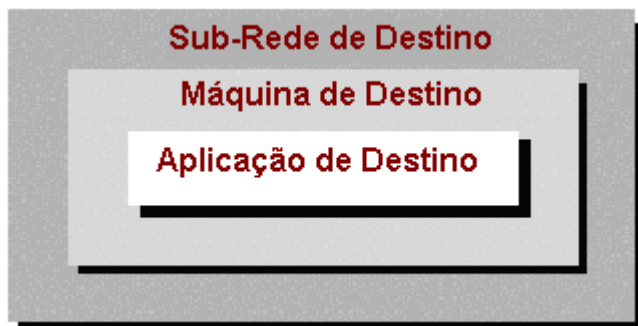



Figura 6-7.2: Classe de equivalência de encaminhamento (FEC)

Rótulo (*Label*):

Um rótulo é um identificador relativamente curto, de tamanho fixo e não estruturado que pode ser usado para ajudar o processo de encaminhamento. Rótulos são associados a uma FEC por um processo de ligação (*binding*). São geralmente locais a um único enlace de dados e não possuem significado global (como o endereço). São análogos ao VPI/VCI usado em ambientes ATM. Como o ATM é uma tecnologia que já usa campos pequenos e de comprimento fixo para tomar decisões de comutação, a comutação por rótulos é avaliada como uma forma efetiva de aplicar IP sobre ATM. Rótulos são ligados a uma FEC como resultado de um evento que indica a necessidade de uma ligação. Esses eventos podem ser divididos em duas categorias:

- Ligação gerada por dados: Ocorrem quando o tráfego começa a fluir, é submetido a um LSR (Roteador de comutação por rótulos) e reconhecido como um candidato a comutação por rótulo. A ligação é estabelecida apenas quando necessária, resultando em menos registros na tabela de encaminhamento. Rótulos são atribuídos a fluxos individuais de tráfego IP e não a pacotes individuais. Em uma rede ATM, isso pode resultar no uso de um número substancial de circuitos virtuais, o que pode limitar a escalabilidade da rede.
- Ligação gerada por controle: São estabelecidas como resultado de uma atividade do plano de controle e são independentes dos dados. Ligações de rótulos podem ser

estabelecidas em resposta a atualizações de roteamento ou recebimento de mensagens RSVP. Esse tipo de ligação é mais facilmente escalável que a abordagem anterior e por esse motivo é usada no MPLS.

6.7.3. Teoria de funcionamento

O roteamento por Rótulo é uma forma avançada de encaminhamento de pacotes que substitui os encaminhamentos convencionais por verificação de endereços longos por um algoritmo de troca de rótulos mais eficiente.

As letras MP (*multi-protocol*) significam que o protocolo pode transportar uma multidão de outros protocolos, isto é: é um protocolo de encapsulamento. As lestras LS (*label switching)* indicam que os protocolos estão sendo encapsulados com um rótulo que é trocado a cada nó.

Existem três distinções importantes entre uma comutação por rótulos e o roteamento convencional:

	Roteamento Convencional	**Comutação por Rótulo**
Analise completa do Cabeçalho IP	Ocorre a cada nó	Ocorre apenas uma vez, na borda da rede, quando o rótulo é atribuído
Suporte a *multicast*	Requer vários algoritmos complexos de encaminhamento	Requer apenas um algoritmo de encaminhamento
Decisões de roteamento	Baseadas apenas no endereço	Podem ser baseadas em qualuqer número de parâmetros, como QoS e VPN.

Tabela 6-7.1: Comutação por rótulos vs. roteamento convencional

Um Roteador de comutação por rótulos (LSR) é um equipamento que suporta tanto o componente de controle padrão IP e componente de encaminhamento por troca de rótulo. A figura abaixo mostra uma rede simplificada de comutação por rótulos e ilustra os LSRs de borda e de interior. Uma rede MPLS serve ao mesmo propósito de qualquer rede de roteamento convencional: entregar tráfego a um ou mais destinos. A adição de encaminhamento por rótulo complementa, mas não substitui o roteamento convencional.

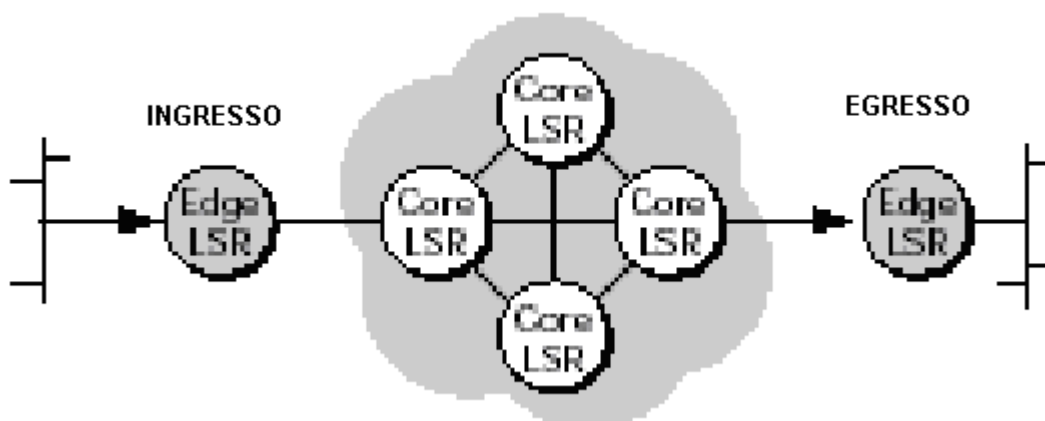



Figura 6-7.3: Rede simplificada de comutação por rótulos

Formato de Rótulos

O grupo de trabalho IETF decidiu que, quando possível, o MPLS deveria usar formatos existentes de rótulos. Por essa razão, MPLS suporta três tipos diferentes de rótulos. Em hardware ATM, usa os bem definidos rótulos VCI e VPI. Em frame relay, usa o rótulo DLCI e em qualquer outro lugar, usa um novo e genérico rótulo conhecido como Shim, que se posiciona entre as camadas 2 e 3. Como o MPLS permite criar novos formatos de rótulos sem ter que trocar os protocolos de roteamento, deve ser relativamente simples estender a tecnologia para formas de transporte óptico emergentes, como DWDM e comutação óptica.

Figura 6-7.4: Formato dos rótulos

O cabeçalho Shim consiste em 32 bits em quatro partes: 20 bits são usados para o

rótulo, 3 bits para funções experimentais, 1 bit para função de empilhamento e oito bits para TTL (*Time to Leave)*.

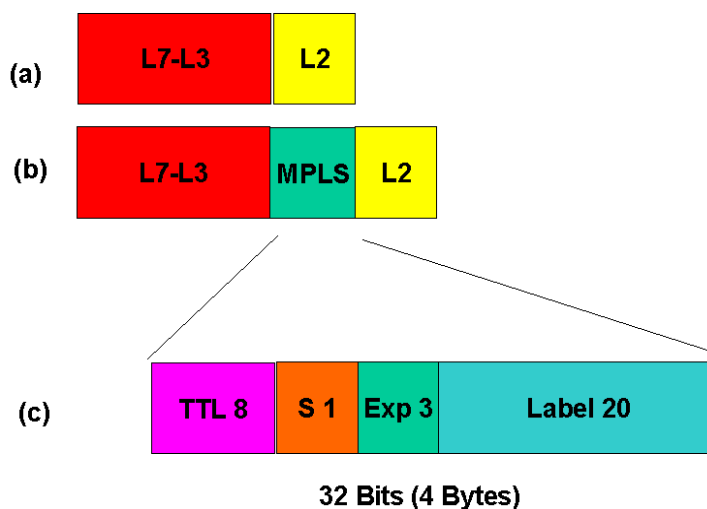



Figura 6-7.5: Cabeçalho Shim

O componente de Encaminhamento no MPLS

Uma vez que uma seqüência de rótulos – chamada de LSP (*Label Switch Path)* ou MPLS tunnel – tenha sido estabelecida, um LSR pode comutar tráfego rapidamente.

Na borda de uma rede MPLS, os LSRs fazem a classificação e decisões de encaminhamento examinando o cabeçalho IP dos pacotes não-rotulados. O resultado é que rótulos apropriados são aplicados aos pacotes e eles são então encaminhados a um LSR que serve como o próximo nó em direção ao destino final.

O rótulo serve como uma representação curta para o cabeçalho IP, reduzindo a complexidade de processamento em todos os nós subseqüentes do caminho. O rótulo é gerado durante o processamento do cabeçalho no nó LSR. Todos os nós subseqüentes na rede usam o rótulo para suas respectivas decisões de encaminhamento. É claro que o valor do rótulo pode mudar (e geralmente muda) a cada LSR no caminho pela rede.

Quando o pacote emerge do interior de uma rede MPLS, os LSRs de borda descobrem que devem encaminhar pacotes a uma interface não-rotulada e simplesmente removem qualquer encapsulamento de pacote antes de encaminhar.

Quando um LSR de interior recebe um pacote rotulado, o rótulo é extraído e usado como índice para a tabela de roteamento residente no LSR. Quando o rótulo extraído é encontrado na tabela, verifica-se o rótulo de saída que é adicionado ao pacote. Este, então, é

enviado à interface de saída encontrada na tabela (se houver multicast, haverá múltiplos pacotes de saída). As tabelas podem ser implementados em nível de nó (uma única tabela por nó) ou em nível de interface (uma tabela por interface).

O mais importante é que apenas um único algoritmo de encaminhamento é necessário para todos os tipos de comutação, e isso pode ser implementado em hardware para maior velocidade.

O MPLS usa FEC, que permite o mapeamento de uma LSP em uma diversidade de formas. Dois pacotes são considerados da mesma FEC se forem colocados na mesma LSP. São suportados mapeamentos por prefixos de endereços IP de tamanho arbitrário ou endereços IP completos de 32 bits. Mapeamentos mais complexos são possíveis como protocolos de controle explicitamente roteados.

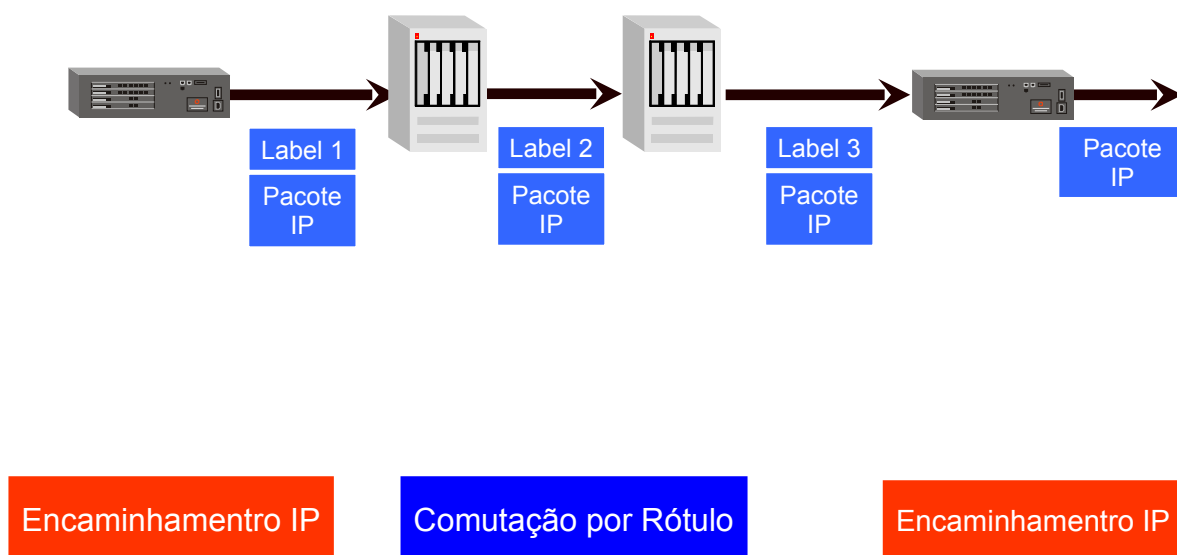



Figura 6-7.6: Fluxo de dados

A forma mais simples de "fluxo de dados" ocorre quando pacotes IP são transmitidos para o LSR de ingresso (LER). Este roteador classifica o tráfego, associando a ele uma FEC. Os LER usam diferentes formas de rotular tráfego. No modelo mais simples, pacotes IP são associados a um rótulo e a uma FEC usando tabelas pré-programadas. Após passar pelo LER, os pacotes são enviados ao próximo LSR, que vai verificar e trocar o rótulo, repassando-o para o próximo LSR:

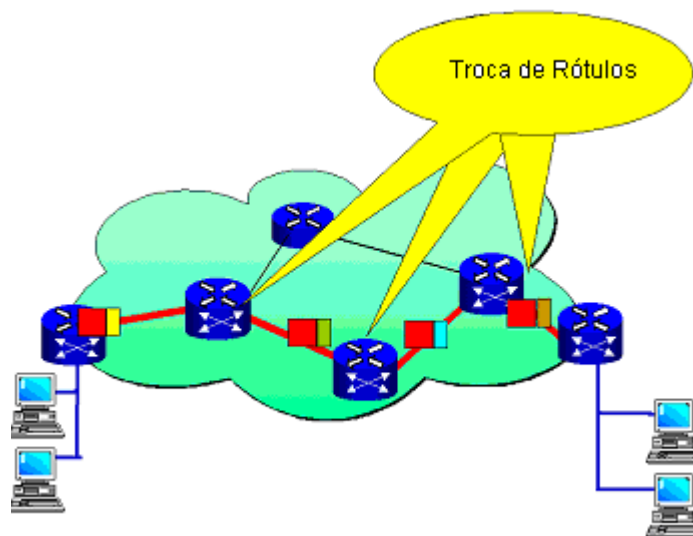



Figura 6-7.7: Troca de rótulos

O componente de Controle do MPLS

Os rótulos são inseridos nos pacotes por um LSR de “subida”. O LSR de “descida”, que recebe esses pacotes rotulados deve saber (ou descobrir) o que fazer com eles. Esta tarefa é de responsabilidade do componente de controle. Para tanto, são usados os componentes de um registro na tabela de encaminhamento como guia.

O estabelecimento e manutenção dessas tabelas são essenciais e devem ser executados por cada LSR. Existem dois modos para carregar as tabelas. Cada roteador poderia ouvir tabelas de roteamento, criar suas próprias tabelas de conexão e informar aos outros a sua informação. Esses roteadores estariam atuando de forma independente. Não haveria nenhum gerenciador de rótulos, e todo roteador teria a possibilidade de ouvir a protocolos de roteamento, gerar tabelas e distribuí-las.

O outro modelo é chamado de *controle ordenado*. Neste caso, um roteador – geralmente o LER de “saída” - é responsável pela distribuição de rótulos.

Cada um dos modelos tem vantagens e desvantagens. Controle independente provê uma rápida convergência. Qualquer roteador que ouvir uma mudança de roteamento pode passar a informação aos outros. A desvantagem é que não há um ponto de controle de tráfego, o que torna a engenharia mais difícil.

O controle ordenado tem a vantagem de melhor engenharia de tráfego e controle mais rígido da rede; entretanto, sua convergência é mais lenta e o controlador é um ponto crítico em termos de falhas.

Dentro do controle ordenado, existem dois grandes métodos de ativar a distribuição de rótulos. São chamados de "down-stream" não solicitado e sob demanda.

Na primeira forma, o gerenciador de rótulos envia-os quando desejar. Ele pode utilizar, por exemplo, um intervalo fixo de tempo. Ou pode utilizar a mudança de tabelas de roteamento padrão como estímulo.

Já na segunda forma, os rótulos são enviados quando pedidos. Podemos ver abaixo que primeiro os rótulos são solicitados para, depois, serem enviados.

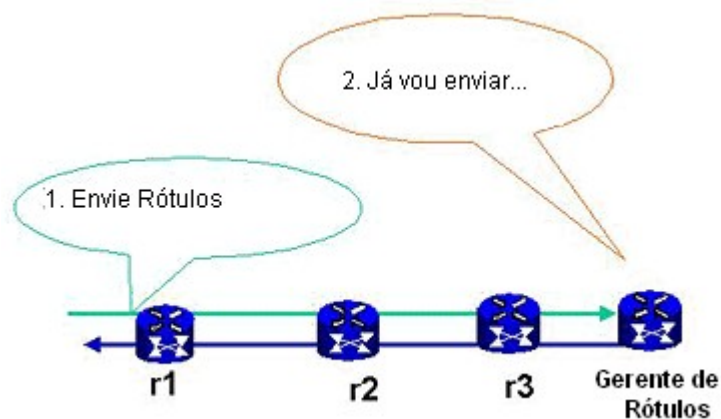



Figura 6-7.8: Gerenciador de rótulos

As tabelas são atualizadas do roteador 3 ao 1, nesta ordem. Os dados contidos são:

- Rótulo de Entrada
- Porta de Entrada
- Rótulo de Saída
- Porda de Saída
- Instrução

No caso mais comum, a instrução é de trocar o rótulo.

Após a atualização das tabelas, os dados podem fluir do roteador de origem ao destino.

Figura 6-7.9: Atualização das tabelas

Protocolos

Encontrar um veículo de transporte para a construção das tabelas é uma grande preocupação dos projetistas de redes. É necessário um protocolo que possa carregar todos os dados necessários e, ao mesmo tempo, ser rápido, auto-recuperável e de alta confiabilidade.

O grupo de trabalho MPLS criou o Protocolo de Distribuição de Rótulos (LDP). Esse protocolo funciona como uma chamada telefônica. Quando os rótulos são associados, permanecem associados até que seja dado um comando para os desfazer. Esse protocolo provê roteamento implícito.

Outros grupos argumentam contra o uso de um protocolo novo e não testado, quando existem protocolos de roteamento que podem ser alterados ou adaptados para carregar as associações. Assim, alguns protocolos já existentes foram alterados para tal função. O Protocolo de Ponte de Borda (BGP) e o IS-IS funcionam bem para tais objetivos.

Os protocolos LDP, BGP e IS-IS estabelecem a LSP, mas fazem pouco pela engenharia de tráfego, pois o tráfego roteado poderia ser redirecionado para uma LSP de maior prioridade, causando congestionamento.

Para ultrapassar esse problema, os protocolos de sinalização foram estabelecidos para criar "túneis" de tráfego (roteamento explícito). Eles são o Protocolo de Distribuição de Rótulos com Rota Restrita (CR-LDP) e o Protocolo de Configuração com Reserva de Recursos (RSVP-TE). Além desses, o Protocolo de Caminho mais Curto (OSFP) foi alterado para suportar engenharia de tráfego (OSPF-TE), mas não é largamente usado.

Protocolo	Roteamento	Engenharia de Tráfego

LDP	Implícito	Não
BGP	Implícito	Não
IS-IS	Implícito	Não
CR-LDP	Explícito	Sim
RSVP-TE	Explícito	Sim
OSPF-TE	Explícito	Sim

Tabela 6-7.2: Protocolos

6.7.4. Garantindo Qualidade de Serviço (QoS)

Para poder disponibilizar uma QoS completa, um sistema deve ser capaz de marcar, classificar e policiar o tráfego. Pode-se entender como uma forma de classificação e marcação, a adição de rótulos, mas a função de policiamento fica faltando. O roteamento e a distribuição de rótulos estabelecem as LSPs, mas ainda não policiam o tráfego e controlam a carga em cada enlace.

Novos componentes de software, que adicionam módulos de gerenciamento entre as funções de roteamento e o seletor de rota, permitem o policiamento e gerenciamento de largura de banda.

Os dois protocolos que fornecem ao MPLS a habilidade de policiar o tráfego e controlar a carga são o RSVP-TE e o CR-LDP.

RSVP-TE

O conceito do processo de estabelecimento de uma chamada, onde os recursos são reservados antes do estabelecimento da mesma, remonta ao passado da sinalização de telefonia. Esse conceito foi adaptado para redes de dados onde a QoS é importante. Um método inicial desenvolvido pelo IETF em 1997, chamado de Protocolo de Reserva de Recursos (RSVP), foi feito para esta função. O protocolo foi criado para solicitar a banda necessária e condições de tráfego necessárias em um caminho definido. Se a banda estivesse disponível nas condições desejadas, o enlace era estabelecido. Para tanto, o tráfego era classificado em três níveis: carga garantida, carga controlada e carga possível.

Este protocolo, adicionado às capacidades para acomodar a engenharia de tráfego MPLS, é chamado de RSVP-TE. Abaixo vemos como um caminho é estabelecido entre duas máquinas. A estação-alvo solicita um caminho específico, com condições de tráfego detalhadas. A mensagem é recebida e uma mensagem de reserva, reservando a banda na rede,

é mandada de volta à estação-alvo. Após a primeira mensagem de reserva ser recebida, os dados podem fluir para um caminho específico de fim a fim.

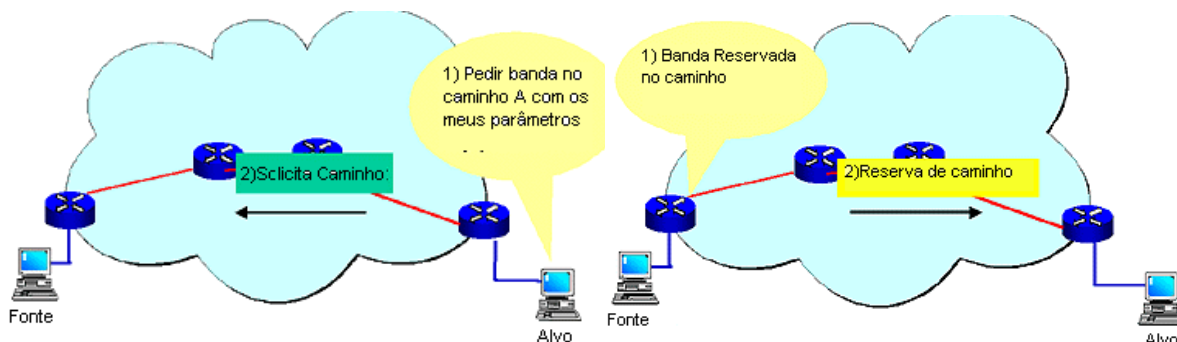



Figura 6-7.10: RSVP-TE

Este tipo de estabelecimento é chamado de "soft state", pois a chamada será derrubada se não houver uma atualização respeitando temporizadores.

CR-LDP

Com o CR-LDP, modificações foram feitas ao protocolo LDP para permitir especificações de tráfego. Esse protocolo adiciona campos ao LDP. Eles são chamados de taxas de dado de pico, comprometido e de excesso, muito similar ao usado em redes ATM.

O estabelecimento de chamada é um processo bem simples: uma requisição e um mapeamento. Diferentemente do protocolo RSVP-TE, o CR-LDP mantém o caminho, uma vez estabelecido, até que uma requisição explícita seja feita.

6.7.5. Vantagens do MPLS

Uma das maiores vantagens do MPLS é o fato de que está se apresentando como uma implementação de comutação por rótulos padronizada. O desenvolvimento de padrões resulta em um ambiente aberto (com vários fabricantes compatíveis). A competição também resulta em preços mais baixos e leva a mais rápidas inovações.

Rotas explícitas

Uma capacidade fundamental do MPLS é o suporte a rotas explícitas. Esse tipo de rota é bem mais eficiente que a opção original do IP. Também provê uma parte da funcionalidade necessária à engenharia de tráfego. Caminhos roteados explicitamente também permitem a criação dos "túneis opacos", que podem levar qualquer tipo de tráfego previamente combinado entre os dois pontos extremos.

Suporte a Multiprotocolo e Mullti-enlace

O componente de encaminhamento MPLS não é específico para nenhuma camada de rede específica. Por exemplo, o mesmo componente de encaminhamento poderia ser usado para IP ou IPX. Também pode operar praticamente sobre qualquer protocolo de enlace, ainda que sua ênfase inicial seja ATM.

Modularidade

Clara separação entre as funções de encaminhamento e controle. Cada parte pode evoluir sem impactar a outra.

Roteamento inter-domínio

Provê uma separação mais completa entre roteamento inter e intra domínio. Isso melhora a escalabilidade dos processos de roteamento e reduz o conhecimento de rotas necessário dentro de um domínio.

Suporte a todos os tipos de tráfego

Outra vantagem menos visível é o suporte a qualquer tipo de encaminhamento: pacote único, único com tipo de serviço e múltiplo.

6.8. VoIP: TRANSMISSÃO DE VOZ SOBRE IP

Ao longo das últimas décadas, pode-se notar um avanço extraordinário das redes de comutação de pacotes em termos de confiabilidade, capacidade e custo. Em contrapartida, pouca coisa evoluiu nas redes de comutação de circuitos, caso da rede telefônica, desde a década de 80.

O maior exemplo de rede que usa a comutação de pacotes para transmissão de dados é a Internet. Ao contrário da comutação de circuitos, na comutação de pacotes o meio é compartilhado, o que permite aumentar a eficiência da utilização dos recursos da rede. A comutação de pacotes é ideal para a transmissão de dados, entretanto ela apresenta alguns problemas para a transmissão de tráfegos que possuem restrições de tempo, como é o caso da transmissão de voz em tempo real.

Em 1999 o número de bytes de tráfego de dados na rede telefônica igualou o tráfego de voz. Já em 2002, o volume do tráfego de dados era superior volume do tráfego de voz. Além disso, o tráfego de dados continuou a crescer exponencialmente, enquanto o tráfego de voz estagnou. Ao observar esta evolução do tráfego, as operadoras telefônicas e os provedores de serviço da Internet se interessaram em transportar voz nas suas redes de dados. Para isso era necessária a criação de técnicas de transmissão e normas para integrar a sinalização e as funcionalidades da rede telefônica e da Internet.

A técnica de transmissão de Voz sobre IP (VoIP, do inglês *Voice over* IP) significa transmitir voz digital pela Internet, utilizando o protocolo de rede IP (*Internet Protocol*). Esta tecnologia é uma alternativa à tradicional transmissão de voz pela rede pública de telecomunicações e tem sido empregada em larga escala. A técnica de voz sobre IP foi proposta pelo VoIP *Forum*, uma iniciativa liderada por grandes fabricantes de equipamentos para promover o uso do padrão ITU-T H.323 na transmissão de vídeo e áudio pela Internet, utilizando o protocolo IP.

As seções seguintes abordam as principais características da transmissão de voz em redes comutadas a pacotes. Será apresentada a motivação para o uso desta técnica, assim como os protocolos usados, os codificadores de voz, os parâmetros de Qualidade de Serviço (QoS - *Quality of Service*) e algumas técnicas para suavizar o efeito das perdas de pacotes.

6.8.1. Motivação

Voz e dados conviveram por muitos anos em redes TDM (*Time Division Multiplexing*), Frame Relay e ATM (*Asynchronous Transfer Mode*). Entretanto, somente com o crescimento acelerado da Internet e, conseqüentemente, o desenvolvimento da tecnologia voz sobre IP foi possível integrar a comunicação destas duas mídias.

A transmissão de voz usando o protocolo IP apresenta uma série de vantagens. Uma das mais significativas é que a tecnologia VoIP independe da camada de enlace para transmitir dados. Ao transmitir voz usando um protocolo da camada 3, no caso o IP, se neutraliza o risco tecnológico das camadas inferiores. Portanto, uma empresa pode investir seguramente em uma tecnologia que funciona com grande parte dos protocolos de LANs e WANs disponíveis hoje e no futuro.

Uma segunda vantagem significativa é o uso de um protocolo de rede único e de uma rede única para o transporte de dados. Com isso, torna-se mais fácil e econômico o desenvolvimento de soluções integradas e de grande valor agregado. Por exemplo, ao atender os requisitos de um cliente que usa voz, páginas web e outros serviços de rede, simultaneamente em um só meio, se obtém uma maior qualidade de comunicação que a rede telefônica por si só não pode igualar, já que só entrega voz.

O custo sem dúvida alguma é uma outra vantagem da telefonia IP. O incentivo ao uso das redes IP, a Internet e/ou as Intranets, como uma solução alternativa às redes telefônicas tradicionais é a base para o desenvolvimento da indústria de voz sobre IP. Estima-se que grande parte das corporações poderá reduzir suas faturas telefônicas mensais à metade, já que ao transportar voz pela Internet, é possível converter todas as chamadas de longa distância em chamadas locais.

Outros aspectos como segurança e confiabilidade também são mais fáceis de se obter em uma rede unificada para transmissão de voz e dados. A escalabilidade também é imediata e linear por usuário, evitando assim as custosas reestruturações nas redes telefônicas, quando estas chegam a sua capacidade máxima.

Atualmente, vemos apenas os primeiros passos da tecnologia voz sobre IP. Sem dúvida, com o avanço da Internet, esta tecnologia ganhará força nos próximos anos e provocará a deterioração da antiga estrutura das empresas do setor telefônico, caso estas não reajam a tempo e em favor da mudança.

6.8.2. Arquitetura

O primeiro passo para transmitir voz em uma rede de comutação de pacotes é a conversão da voz analógica para a forma digital. Esta conversão é feita através do uso de codificadores de voz, figura 6-8.1.

Figura 6-8.1: Etapas da transmissão de voz sobre IP

A etapa seguinte é a geração do pacote. A informação binária contida na saída do codificador de voz é armazenada até que se consiga atingir o tamanho estipulado para a parte de dados do pacote. Em seguida, acrescentam-se os cabeçalhos necessários.

Cada pacote de voz gerado é transmitido nó-a-nó até o seu destino. Por fim no destino, os pacotes são recebidos, a informação é decodificada e a voz é reproduzida. A figura 6-8.2 mostra a arquitetura voz sobre IP.

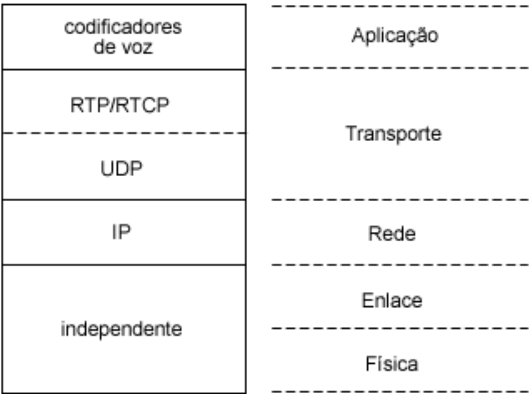



Figura 6-8.2: A arquitetura voz sobre IP

Pode-se perceber que existe uma determinada sobrecarga referente ao acréscimo dos cabeçalhos dos protocolos, isto é, para cada pacote de voz devem ser acrescentados 12 bytes referentes ao cabeçalho do protocolo RTP (*Real-Time Protocol*), 8 bytes referentes ao protocolo UDP (*User Datagram Protocol*), 20 bytes referentes ao protocolo IP. No total, são 40 bytes, além do cabeçalho da camada de enlace.

Do ponto de vista de sobrecarga de cabeçalhos e de processamento dos protocolos, deve-se enviar a maior quantidade possível de informação de voz em cada pacote para maximizar a utilização da capacidade da rede. No entanto, quanto maior a informação de voz, maior o tempo de espera para a geração do pacote e maior o tempo de transferência nó-a-nó na rede de comutação por pacotes. Portanto, existe um compromisso entre a eficiência e o atraso.

6.8.3. Protocolos

Os sistemas de voz sobre IP utilizam os protocolos da arquitetura TCP/IP como infra-estrutura para os seus protocolos de aplicação. A figura 6-8.3 mostra a estrutura em camadas dos principais protocolos usados nos sistemas VoIP.

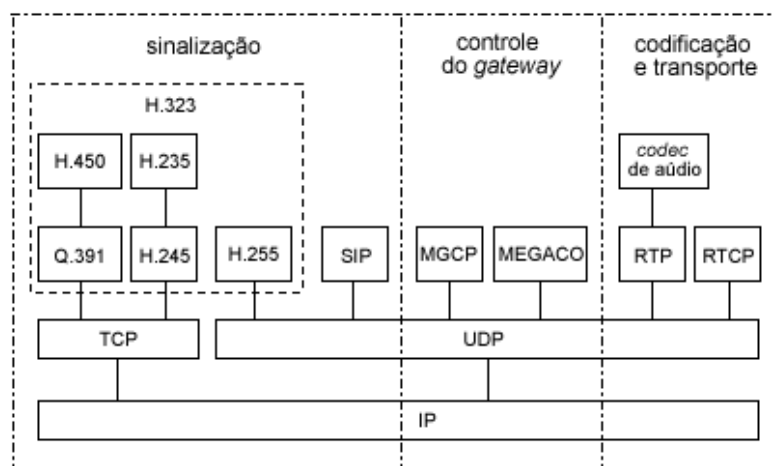



Figura 6-8.3: Protocolos e a arquitetura em camadas para sistemas VoIP

As seções seguintes apresentam alguns dos protocolos de transporte, controle e sinalização utilizados em sistemas de voz sobre IP.

RTP (*Real-Time Protocol*)

O protocolo RTP provê serviços fim-a-fim para aplicações de tráfego em tempo-real. Por isso, ele é usado para transportar pacotes de voz. As principais funcionalidades oferecidas pelo RTP são: a identificação do tipo de tráfego, o número de seqüência de pacotes, a etiqueta de temporização (*timestamping*) e, com o auxílio do RTCP (*Real-Time Control Protocol*), o monitoramento da entrega dos pacotes.

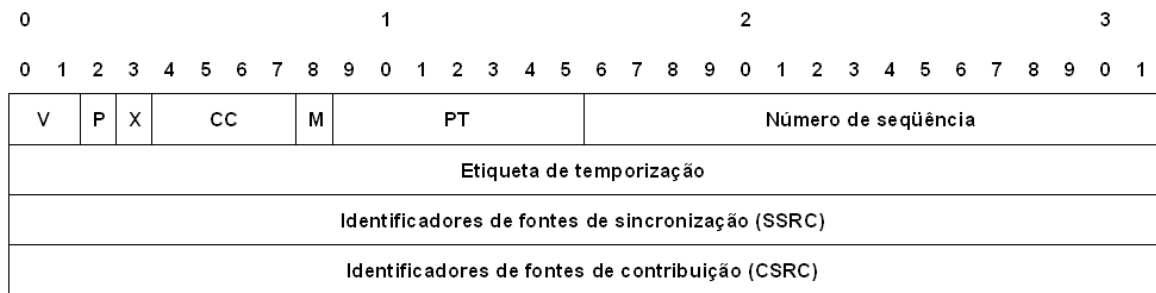



Figura 6-8.4: O cabeçalho do RTP

O cabeçalho do RTP é composto por 12 bytes, figura 6-8.4. Os dois primeiros bits indicam a versão do protocolo. O bit P (*padding*) sinaliza se houve ou não preenchimento dos dados (*payload*) para fins de alinhamento. O último bit do *payload* contém o número de bytes acrescentados ao *payload* original. Já o bit X indica a existência de extensões no protocolo entre o cabeçalho e os dados. O campo CC é preenchido com o número de identificadores de fontes contribuintes, que vem após o cabeçalho fixo e pode variar de 0 a 15. O bit M pode ser usado pela aplicação para marcar determinados pacotes. O tipo de tráfego transportado pelo RTP é identificado pelo campo PT. O próximo campo indica o número de seqüência, que é iniciado aleatoriamente e é incrementado a cada pacote RTP enviado. A etiqueta de temporização determina o tempo entre o envio de pacotes RTP (*interpacket gap*). O campo "identificadores de fontes de sincronização" contém a identificação de um participante dentro de uma sessão RTP. Esse identificador, escolhido de forma aleatória, está diretamente associado à mídia e ao relógio utilizado para gerar as informações. Finalmente, o campo "identificadores de fontes de contribuição" indica as fontes que contribuíram com dados para a formação do pacote. Estes identificadores são os próprios identificadores de sincronismo das fontes contribuintes, que são repassados para a fonte de sincronismo atual.

Usualmente, o RTP é utilizado sobre o protocolo UDP (*User Datagram Protocol*), que provê um serviço de transporte não orientado a conexão (não se envia recibos de mensagens). Isto porque em uma transmissão em tempo real, a retransmissão de pacotes perdidos em geral é inútil, sendo preferível a perda de pacotes à recepção de pacotes atrasados. Além disso, o RTP ainda faz uso da multiplexação e do checksum providos pelo UDP, isto é, o envio de pacotes por diferentes caminhos e o controle de erro, respectivamente. A reordenação dos pacotes fica a cargo do receptor, através do número de seqüência dos pacotes. Deve-se ressaltar que, caso um pacote sofra um atraso acima de um valor limite, ele é considerado perdido pelo receptor.

RTCP (*Real Time Control Protocol*)

O RTCP é um protocolo de controle utilizado em conjunto com o RTP, responsável por monitorar a qualidade do serviço e por repassar informações sobre participantes de uma dada sessão RTP. Uma sessão RTP consiste em um conjunto de participantes que se comunicam através do protocolo RTP. Para cada sessão são necessárias duas portas de comunicação UDP diferentes: uma utilizada pelo RTP e outra pelo RTCP.

As informações obtidas pelo RTCP estão relacionadas a características da sessão, tais como: os participantes, a variação do atraso, a taxa de perdas, entre outras.

H.323

A norma H.323, publicada pelo ITU-T (*International Telecommunication Union – Telecommunication Standardization Sector*) em 1996, define um padrão para ser utilizado em conferências através da Internet. Para viabilizar o estabelecimento de chamadas telefônicas na Internet, o H.323 especifica os protocolos de transporte, de controle e de sinalização, bem como os codificadores de voz. Por isso é conhecido como um protocolo "guarda-chuva", pois define todos os requisitos necessários para a transmissão de voz sobre IP, desde a geração do pacote de voz até o estabelecimento da chamada e a da troca de disponibilidade de recursos da rede. O padrão H.323 não provê nenhum tipo de garantia de qualidade de serviço.

Podemos citar alguns benefícios da adoção do padrão H.323:

- **Independência da rede**

O padrão H.323 é projetado para utilização em redes baseadas em pacotes, como as redes IP. Na atualidade, a maioria das redes possui uma infra-estrutura com protocolo de transporte baseado em pacotes, assim a adoção do padrão H.323 permite a utilização de aplicações multimídia sem requerer mudanças na estrutura das redes

- **Interoperabilidade de equipamentos e aplicações**

O H.323 permite interoperabilidade entre dispositivos e aplicações de diferentes fabricantes. Por isso, vários fornecedores de porte como Intel, Microsoft, Cisco e IBM investem em linhas de produtos H.323.

- **Independência de plataforma**

O H.323 não define o hardware ou sistema operacional a ser usado. Desse modo, as aplicações H.323 podem ser de naturezas diversas voltadas para mercados específicos, que vão desde software de videoconferência executado em PCs, a telefones IP, adaptadores para TV a cabo, sistemas dedicados, etc.

- **Representação padronizada de mídia**

O H.323 estabelece codificadores para compressão e descompressão de sinais de áudio e vídeo. Ele também prevê mecanismos de negociação dos codificadores a serem utilizados numa conferência a fim de que os seus participantes encontrem um subconjunto comum entre si.

- **Flexibilidade nas aplicações clientes**

Uma conferência H.323 pode envolver aplicações clientes com capacitações multimídia diferentes. É possível que um terminal com suporte apenas para áudio participe de uma conferência com terminais que tenham suporte adicional de vídeo e/ou dados.

- **Interoperabilidade entre redes**

É possível estabelecer conferências entre participantes localizados numa LAN e em outras redes completamente diferentes, como a rede telefônica pública ou ISDN. O H.323 prevê o uso de codificadores que são comuns a vários tipos de redes. Isto é possível através da utilização do componente *gateway*.

- **Suporte a gerenciamento de largura de banda**

O tráfego dos fluxos de vídeo e áudio é caracteristicamente consumidor de largura de banda em uma rede. O padrão provê mecanismos de gerenciamento que permitem delimitar a quantidade de conferências simultâneas e a largura de banda destinada às aplicações H.323. Além do mais, o H.323 também prevê facilidade de contabilidade de uso dos recursos da rede que podem ser usadas para fins de cobrança. Isto é possível através do uso de um *gatekeeper*.

- **Suporte a conferências multiponto**

O H.323 suporta conferências com três ou mais participantes simultâneos.

- **Suporte a *multicast***

O H.323 suporta técnicas de *multicast* nas conferências multiponto. Uma mensagem *multicast* envia um único pacote a todo um subconjunto de destinatários na rede sem replicação. Esse tipo de transmissão usa a largura de banda de uma forma muito mais eficiente que as transmissões *unicast*.

A figura 6-8.5 mostra a arquitetura H.323 para a telefonia IP. A norma H.323 permite a realização de chamadas dentro da Internet, da Internet para e rede telefônica e da rede telefônica para a Internet.

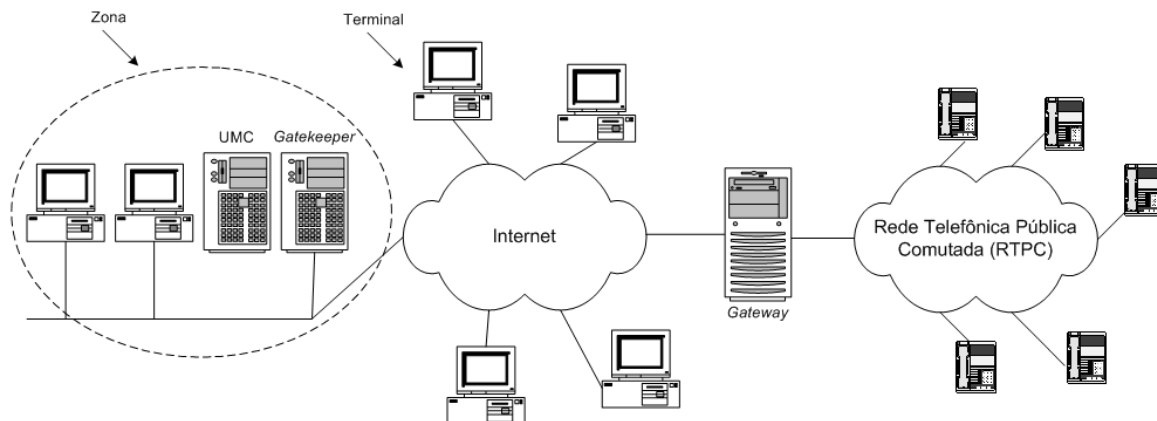



Figura 6-8.5: A arquitetura H.323

As duas redes são interligadas por um dispositivo de passagem denominado *gateway*. A função do *gateway* é prover a interoperabilidade entre a Internet e a Rede Telefônica Pública Comutada (RTPC), já que estas redes utilizam protocolos diferentes. De forma simplificada, a função do *gateway* é permitir a comunicação entre terminais H.323 e outros tipos de terminais. Ele executa a conversão da voz analógica em voz digital comprimida em tempo real, ou vice versa, e a conversão de sinalização para as chamadas telefônicas.

A Unidade de Controle Multiponto (UCM), ou MCU é responsável pelos serviços de conferência entre três ou mais terminais. É composta por um Controlador Multiponto (CM), responsável pela sinalização das chamadas, e por um Processador Multiponto (PM), responsável pelo processamento dos pacotes de dados dos sinais de voz dos terminais envolvidos na conferência. Ela pode operar em três modos distintos: Centralizado, descentralizado ou híbrido.

- **Modo Centralizado**: A Comunicação entre a MCU e os terminais ou Gateway é unicast. Dados, áudio, vídeo e controle passam obrigatoriamente pela MCU.

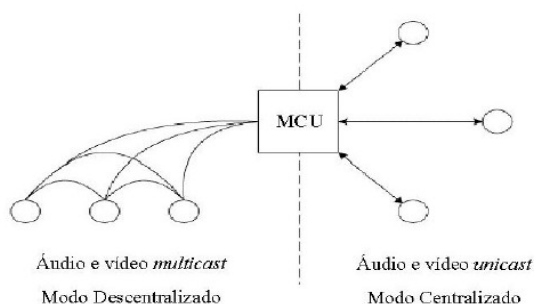



Figura 6-8.6: Modo Centralizado

- **Modo Descentralizado**: Os terminais trocam informações de controle e opcionalmente de dados de forma centralizada com a MCU, mas trocam áudio e vídeo entre si por multicast.

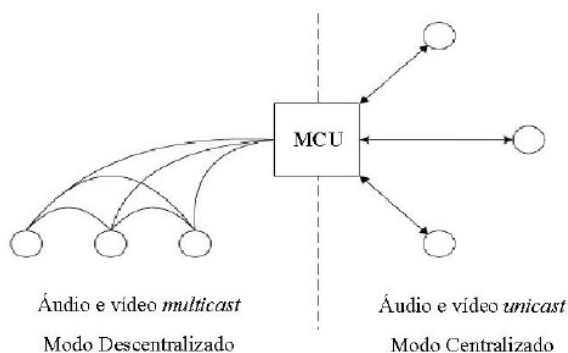



Figura 6-8.7: Modo descentralizado

- **Modo Híbrido**: A comunicação de dados e controle sempre se dá de forma centralizada com a MCU. Contudo podemos ter áudio também centralizado e multicast de vídeo ou vice-versa.

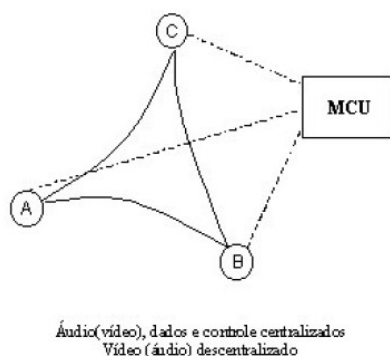



Figura 6-8.8: Modo híbrido

O *gatekeeper* é o equipamento responsável pelo gerenciamento de um conjunto de equipamentos dedicados à telefonia IP. Suas principais funções são: executar a tradução de endereçamento dos diversos equipamentos, controlar o acesso dos equipamentos à rede dentro de sua zona, e controlar a banda utilizada. É um elemento opcional no ambiente H.323. Podemos destacar outras funcionalidades para ele:

- Controle de banda – rege os pedidos de troca de banda em uso;
- Pode controlar toda a sinalização entre chamadas de terminais;

Uma zona é um conjunto de terminais, *gateways* e UCMs gerenciados por um único *gatekeeper*. Uma zona deve ter pelo menos um terminal, e pode ou não conter *gateways* ou UCMs. Entretanto, uma zona tem apenas um *gatekeeper*. Fisicamente, a zona pode ser composta por um ou mais segmentos de rede interligados através de roteadores ou outros equipamentos semelhantes. Comparada com os sistemas telefônicos convencionais, uma zona corresponde a uma área com um determinado código de localidade.

A figura 6-8.9 resume o escopo da recomendação H.323:

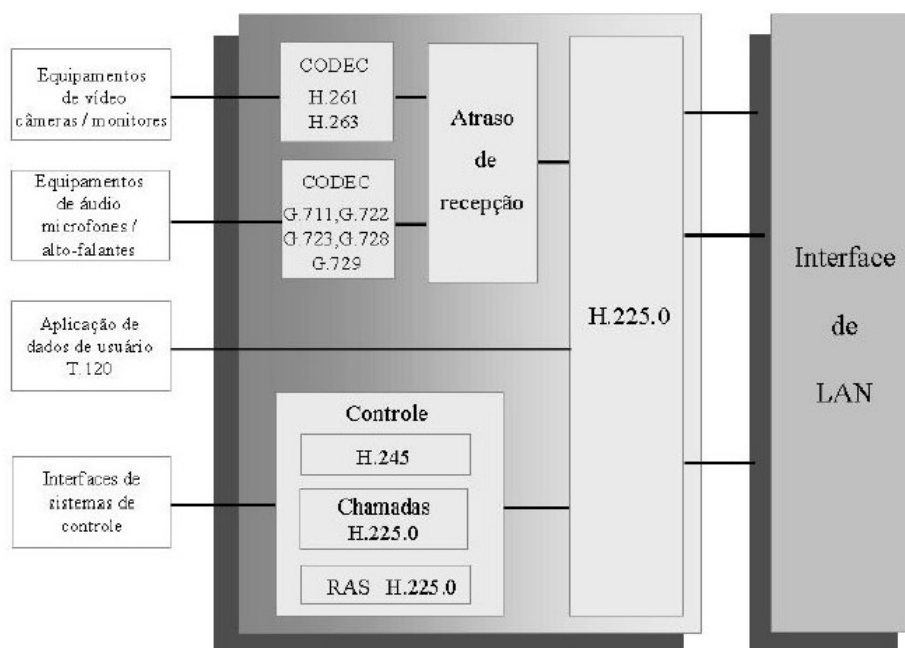



Figura 6-8.9: Recomendações H.323

Como visto anteriormente, a norma H.323 especifica os requisitos e protocolos para sistemas de transmissão de voz em redes baseadas em pacotes. Para a codificação e decodificação da voz, um dos padrões sugeridos é o G.711, suportado por todos os terminas H.323. Outros codificadores, com maior taxa de compressão, podem ser usados como será mostrado adiante.

O padrão H.323 recomenda o uso tanto do RTP quanto do RTCP para o transporte e controle da qualidade do tráfego de voz. Já os protocolos H.245 e Q.391 são responsáveis, respectivamente pela negociação do protocolo de codificação e pelo o estabelecimento e o fechamento das ligações. Por fim, o protocolo H.225 é o responsável pela comunicação dos terminais com o *gatekeeper*. Ele gerencia o canal de comunicação, chamado de RAS (*Registration, Admission, Status*) entre os dois dispositivos. O H.225 permite a união de

terminais, a movimentação de um terminal para fora da sua zona e a negociação da largura de banda. Podemos observar os protocolos e suas respectivas camadas usados no H.323 na figura 6-8.10:

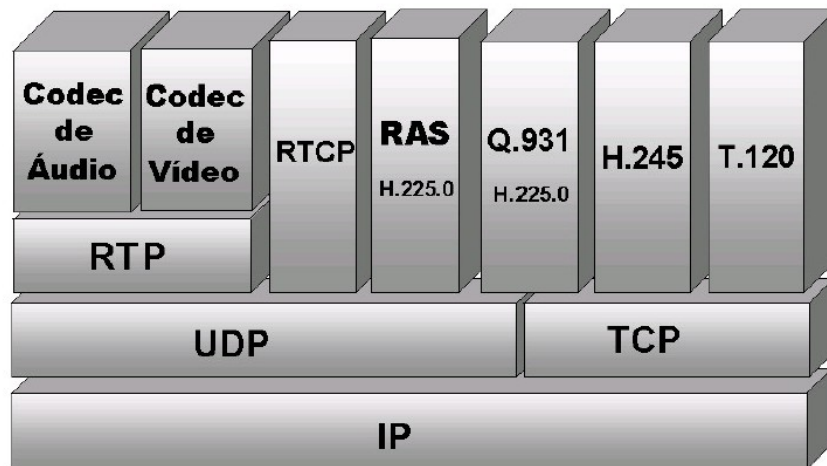



Figura 6-8.10: Protocolos utilizados no ambiente H.323

Uma chamada a partir de um terminal H.323 em uma rede local com *gatekeeper* para um telefone remoto é exemplificada na figura 6-8.11.

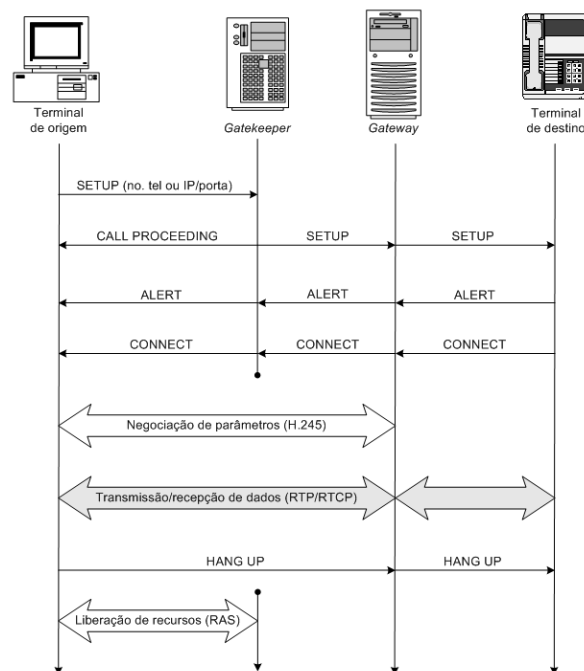



Figura 6-8.11: Exemplo de uma chamada de um terminal H.323 para um telefone remoto

O terminal H.323 difunde pela rede um pacote UDP para descobrir o endereço IP do *gatekeeper*. Conhecendo o endereço do *gatekeeper*, o terminal envia uma mensagem de registro RAS para este. Depois de ter o seu pedido de registro aceito pelo *gatekeeper*, o terminal envia mensagens RAS de admissão e negociação da largura de banda. Se esta negociação for bem sucedida, um canal de comunicação entre o terminal e o *gatekeeper* é estabelecido para a troca de mensagens de sinalização Q.391 de forma a criar a ligação com o telefone remoto.

Após ser estabelecida a ligação telefônica, o terminal H.323 passa a se comunicar diretamente com o *gateway*. Durante a negociação de parâmetros, usando o protocolo H.245, cada dispositivo envolvido anuncia os seus recursos, como chamadas em conferência, codificadores suportados, etc. Depois, são estabelecidas duas conexões unidirecionais RTP, eventualmente com codificações diferentes em cada sentido. A sincronização de áudio é realizada com RTCP. Durante uma chamada estarão abertos cinco canais de comunicação entre os terminais: um canal de sinalização da chamada Q.931, um canal de controle da chamada H.245, dois canais RTP, um direto e outro reverso, de envio de dados e um canal RTCP de controle dos dados.

Qualquer um dos terminais pode terminar a ligação, usando o sinal HANGUP do protocolo de sinalização Q.931. Depois de terminada a ligação, o terminal H.323 informa ao *gatekeeper*, que os recursos alocados naquela chamada já podem ser liberados.

SIP (*Session Initiation Protocol*)

O padrão H.323 é visto pela comunidade da Internet como um produto típico das operadoras telefônicas: extenso, complexo e inflexível. O SIP foi o protocolo proposto pela IETF para controlar o estabelecimento de chamadas telefônicas, de videoconferências, e outras ligações multimídias.

O SIP é um módulo isolado que lida apenas com o estabelecimento de ligações. Estas ligações podem ser ponto-a-ponto, multiponto e ponto-a-multiponto. Além disso, o SIP provê os serviços de localização de um terminal, de determinação dos recursos de um terminal e de sinalização para estabelecimento e encerramento de chamadas.

Num sistema SIP, o endereço de um terminal é dado por um URL (*Uniform Resource Locator*) que pode conter endereços IP, versão 4 ou 6, ou números de telefone.

O SIP possui uma estrutura semelhante a do protocolo HTTP (*Hyper Text Tranfer Protocol*) tipo cliente-servidor. Requisições são geradas pelo cliente e enviadas ao servidor

(entidade receptora) que processa o pedido e envia a resposta novamente ao cliente. As mensagens trocadas são formadas por caracteres ASCII. Os pedidos são compostos pelo nome de um método, seguido de linhas adicionais contendo parâmetros. As respostas incluem um código de três dígitos e também podem ser seguidas de várias linhas adicionais contendo parâmetros. A tabela 6-8.1 contém alguns dos métodos usados pelo protocolo SIP.

O estabelecimento de uma ligação é realizado através do reconhecimento em três vias (*three-way-handshake*). O terminal que está solicitando a chamada envia uma mensagem INVITE para o terminal de destino. Esta solicitação pode ser enviada em cima do TCP (*Transmission Control Protocol*) ou do UDP. Caso aceite a solicitação, o terminal destino responde com OK e com a lista de recursos suportados. Para estabelecer a conexão, o terminal solicitante envia um ACK ao receber a resposta do INVITE. Qualquer um dos terminais pode terminar a ligação com o envio de uma mensagem BYE.

Método	**Descrição**
INVITE	Pedido de início de sessão.
ACK	Confirmação de início de sessão.
BYE	Pedido de terminação da sessão.
OPTIONS	Informação sobre os recursos do terminal.
CANCEL	Cancelamento de um pedido pendente.
REGISTER	Informa a um servidor de redirecionamento a posição atual de um usuário.

Tabela 6-8.1: Métodos do SIP

Como mencionado anteriormente, as respostas do protocolo aos pedidos acima são respostas de três dígitos em que o primeiro dígito especifica o tipo de mensagem e os outros dois o código da mensagem. São os diferentes tipos de resposta:

- 1XX – Progresso
- 2XX – Requisição bem sucedida
- 3XX – Redirecionamento
- 4XX – Requisição Incorreta
- 5XX – Falha do servidor
- 6XX – Falha Global

O ambiente SIP possui três servidores diferentes para os seguintes casos:

- Servidor de Registro: recebe atualizações da localização dos usuários na rede.
- Servidor Proxy: recebe requisições e encaminha para outros servidores que contenham informações mais precisas sobre a localização do usuário.
- Servidor de Redirecionamento: recebe requisições. Ao invés de as encaminhar para servidores retorna o endereço do servidor para o qual a requisição deve ser direcionada.

Dessa forma a figura 6-8.12 exemplifica um pedido de conexão utilizando os três tipos de servidores existentes.

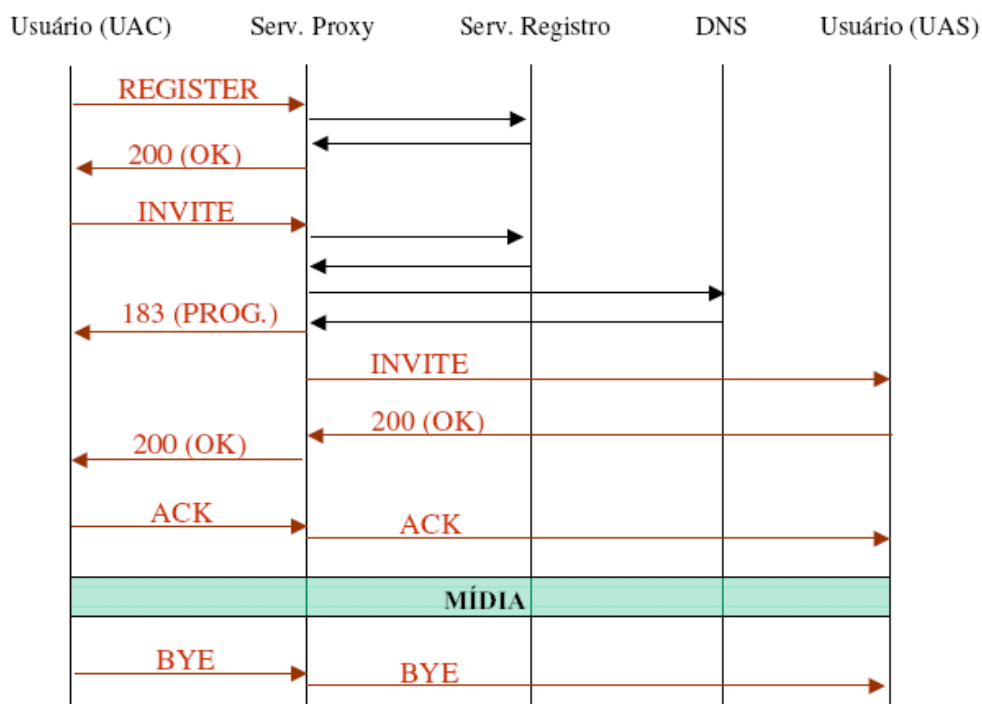



Figura 6-8.12: Estabelecimento de chamada com redirecionamento.

Inicialmente o usuário cliente (UAC) faz o pedido de registro ao servidor Proxy que envia ao servidor de registro a informação sobre sua localização. A confirmação (200) vem em seguida garantindo o sucesso da operação. Em seguida este usuário já envia um pedido de conexão que passa pelos três servidores até ser estabelecida a chamada. Neste caso o pedido de Invite foi redirecionado para chegar no usuário cliente.

Para lidar com usuários móveis, isto é, usuários que podem migrar na rede, o SIP define a utilização de um *proxy*, que torna transparente a localização do usuário para os outros

terminais. A informação de localização de usuário é mantida com o uso de um servidor de localização. Cada vez que muda de localização, o usuário envia um REGISTER para atualizar a sua localização. Após receber o INVITE, o *proxy* pesquisa o servidor de localização, usando um protocolo externo ao SIP, para obter a localização do usuário. Em seguida, o *proxy* serve de intermediário reenviando as mensagens recebidas dos dois participantes, até ao estabelecimento da ligação. Esta situação é ilustrada na figura 6-8.13.

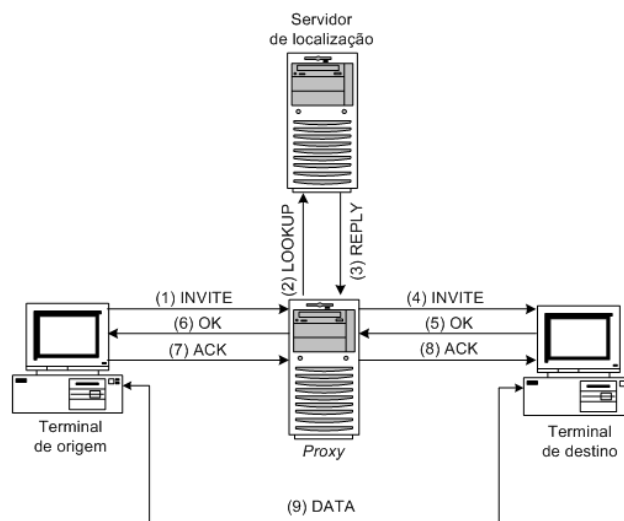



Figura 6-8.13: O papel do servidor de localização e do *proxy*

O protocolo SIP não recomenda o uso de nenhum protocolo para o transporte e o controle da qualidade do tráfego de voz. Entretanto, costuma-se usar os protocolos RTP e RTCP.

Essencialmente o SIP foi desenvolvido para a realização de chamadas dentro da Internet. No entanto, utilizando *gateways* é possível realizar chamadas para a RTPC. O SIP ainda suporta a espera de chamadas, a encriptação e a autenticação.

6.8.4. Codificação da Voz

A função dos codificadores de voz é converter a voz da forma analógica para a forma digital. A codificação da voz consiste na amostragem e na quantização do sinal. A amostragem transforma o sinal contínuo em um sinal discreto, enquanto a quantização converte o sinal discreto analógico em uma seqüência de bits.

A faixa de freqüência da voz humana utilizada para telefonia é de 4,0 kHz, e segundo a taxa de Nyquist, a voz deve ser amostrada pelo menos no dobro da freqüência para que não haja *aliasing*. Assim, têm-se 8 mil amostras por segundo. A figura 684.1 ilustra esta situação

Existem três tipos básicos de codificadores de voz: formato de onda, paramétrico e híbrido. O codificador por formato de onda procura transformar em bits as características da forma de onda do sinal de fala, um exemplo deste tipo é o codificador PCM. Já o codec paramétrico tem como objetivo uma maior compactação do sinal de voz e por este motivo realiza uma modelagem deste sinal e em seguida codifica seus parâmetros. Por fim, o codificador híbrido realiza uma mistura entre os outros dois tipos, codificando tanto as características da forma de onda do sinal de voz com seus parâmetros do modelo. A Figura 6-8.14 mostra uma comparação dos tipos de codecs levando em consideração a qualidade da voz e sua taxa de bits.

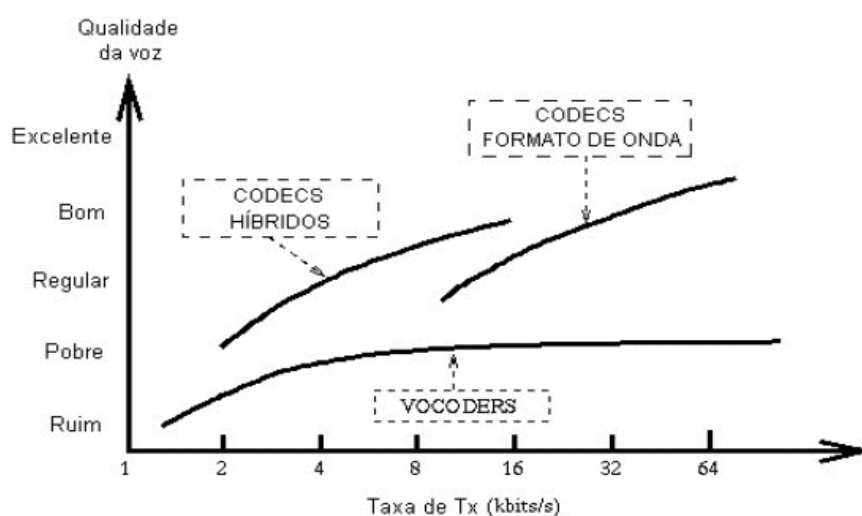



Figura 6-8.14: comparação entre os diferentes tipos de codecs.

Podemos perceber então que os codificadores por formato de onda são os que melhor representam a voz, mantendo assim uma qualidade mais alta o que implica em uma taxa de bits elevada (64kbps). Entretanto os codificadores híbridos, quando analisados para uma qualidade de voz de boa para regular, mostraram-se mais eficientes que os codecs por formato de onda, apresentando uma taxa de bits mais baixa.

O G.711 é o codificador mais básico definido pela ITU-T. Este codificador utiliza a modulação PCM (*Pulse Code Modulation*) em que cada amostra é representada por um byte. Sendo assim, a taxa de transmissão da voz PCM é de 64 kbps. Na quantização é utilizada uma escala logarítmica de modo a aumentar a resolução de sinais com níveis de tensão mais baixos, onde se encontra a maior parte da informação do sinal. O algoritmo PCM é um codificador simples por formato de onda e insere um atraso de 0,125ms, que é o tempo de

codificação de apenas 1 bit. A escala logarítmica de conversão está ilustrada na Figura 6-8.15.

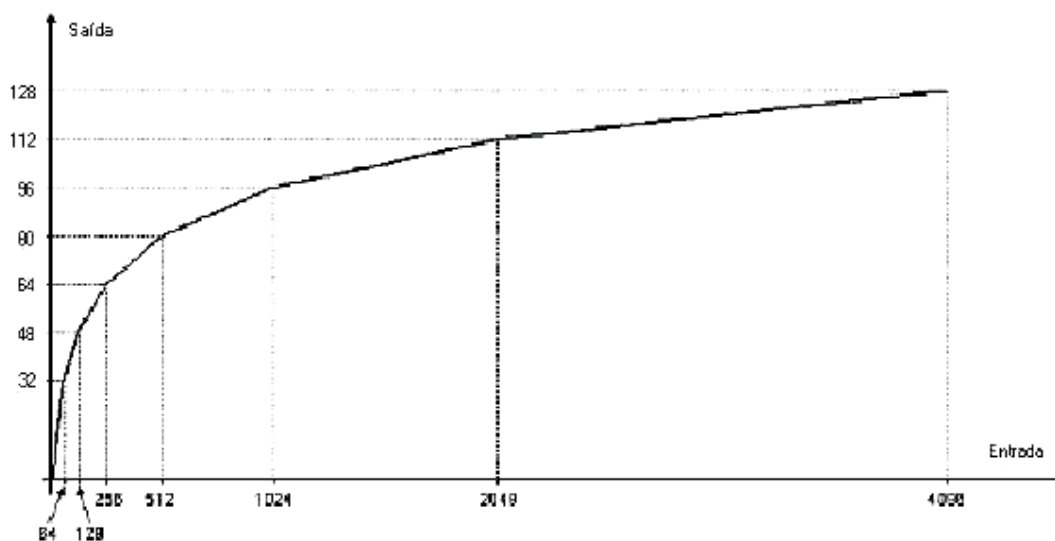



Figura 6-8.15: Níveis de entrada e saída na quantização do sinal de voz

Existem codificadores de voz que utilizam a técnica de supressão de silêncio. Esta técnica se baseia no fato de que durante uma conversação o interlocutor fala em média apenas 35% do tempo. Para suprimir o silêncio, é necessária a existência de um detector de voz (*Voice Activity Detector* - VAD), responsável por detectar se o interlocutor está falando ou está em silêncio. É necessário também um dispositivo, o DTX (*Discontinuous Transmission*), que interromperá a transmissão toda vez que receber a informação de que o interlocutor está em silêncio. Por fim, é necessário um outro dispositivo, o CNG (*Confort Noise Generator*), para simular o ruído de fundo enquanto o interlocutor estiver em silêncio, pois o silêncio absoluto causa desconforto aos ouvidos humanos.

Existem ainda diversos outros codificadores de voz que possuem taxas de transmissão inferiores. A grande maioria destes codificadores faz a compressão do sinal de voz em quadros, onde cada quadro contém diversas amostras de voz. Estas técnicas de compressão utilizam informações dos quadros anteriores para a geração dos quadros atuais. Estes codificadores possuem também em memória dicionários com quadros (formas de onda pré-gravadas) mais prováveis, de forma que o algoritmo verifica qual desses quadros se assimila mais ao gerado. Além disso, ajusta um conjunto de filtros e ganhos para que o sinal fique bem próximo ao sinal real o que caracteriza uma codificação paramétrica. Um esquema desta implementação segue na Figura 6-8.16.

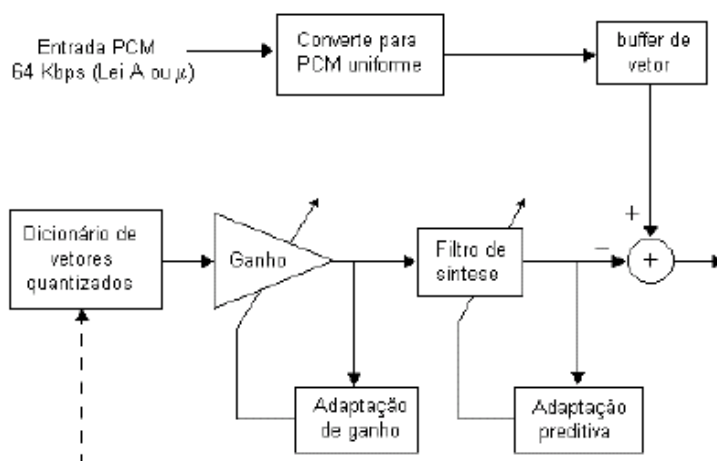



Figura 6-8.16 Esquema básico de um codificador híbrido com dicionário de formas de onda, filtros e amplificadores.

Estes codecs são então chamados de codificadores híbridos, pois misturam características sobre a forma de onda, nos dicionários, e os parâmetros de ajuste de filtros e amplificadores para modelar o sinal a ser codificado. Dentre estes codificadores citamos G.729 e G723.1, que são os codificadores mais utilizados atualmente no ambiente H323.

O codificador G.729A utiliza a técnica de codificação CS-ACELP (*Conjugate Structure Algebraic-Code-Excited Linear Prediction*). Ele possui uma taxa de 8 kbps com tamanho de quadro de 10 ms, ou seja 80 amostras, mais 5 ms de informações do quadro anterior (*look-ahead*). Assim, o atraso do algoritmo é de 15ms.

O codificador G.723.1 foi selecionado pela *International Multimedia Telecommunications Consortium* (MTC) VoIP *Forum* para o padrão H.323 banda estreita. Este codificador possui uma taxa de transmissão de 5,3 ou 6,3 kbps, que pode ser ajustada dinamicamente. O tamanho do quadro é de 30 ms mais 7,5 ms de informações do quadro anterior. Cada quadro é dividido em conjuntos de 60 amostras de modo que cada 240 amostras formam 30ms (6 quadros). O G.723.1 utiliza o MP-MLQ (*Multi-Pulse-Maximum Likelyhood Quantization*) para excitação da taxa mais alta, enquanto que para a taxa mais baixa é utilizado o ACELP (*Algebraic-Code-Excited Linear Prediction*). Tanto o G.723.1 quanto o G.729A são capazes de realizar a supressão de silêncio.

Codificador	Taxa	Quadro	MOS
G.711 (PCM)	64 kbps	1 ms	4,1
G.723.1 (MP-MLQ)	6,3 kbps	30 ms	3,9

G.723.1 (ACELP)	5,3 kbps	30 ms	3,65
G.729A (CS-ACELP)	8,0 kbps	10 ms	3,7

Tabela 6-8.2: Codificadores de voz

A tabela 6-8.2 apresenta as principais características dos codificadores anteriormente mencionados e o valor da escala MOS (*Mean Opinion Score*) para cada um. Esta escala determina a qualidade de um sinal de voz. É um parâmetro bem subjetivo, já que é determinado por ouvintes, que atribuem valores entre 0 e 5 ao sinal de voz de acordo com a sua opinião sobre a qualidade do sinal.

O Codec ILBC foi um padrão lançado em dezembro de 2004 pelo IETF com o objetivo de ser utilizado em ambientes de Internet e SIP. Este codificador é híbrido e possui taxas de 13.3 e 15.3 kbps, com atrasos de 30 e 20ms, respectivamente. Apesar de suas taxas de bits superarem as do G.723.1 e G.729A, este codificador se mostrou bem mais robusto que os dois últimos em ambientes com perdas, que é o caso de comunicação VoIP na Internet.

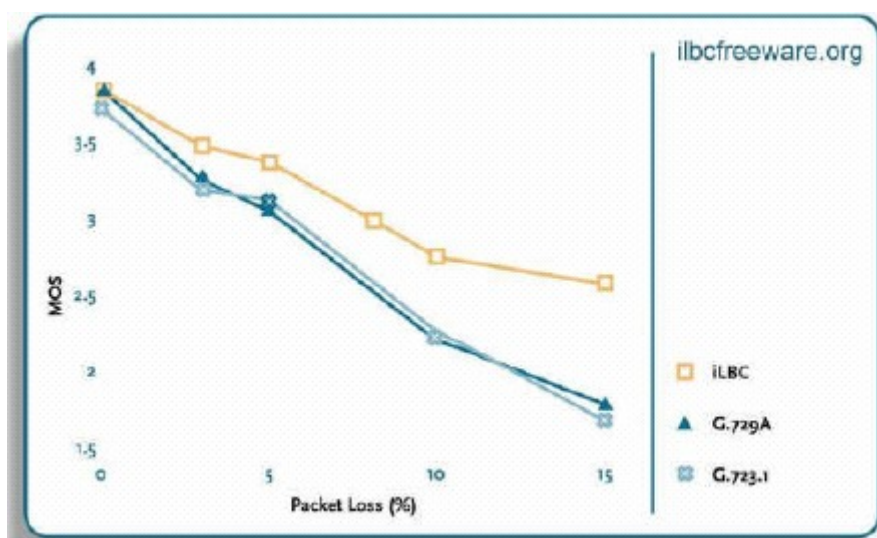



Figura 6-8.17: Comparação entre o codec ILBC e o G.723.1 e G.729A para ambientes com perdas.

A Figura 6-8.17 mostra um gráfico da qualidade da voz em função do aumento de perdas comprovando o fato do codec ILBC ser mais robusto em ambientes críticos. Isto ocorre devido ao fato de que os quadros gerados por este codec serem independentes entre si, não necessitando do quadro anterior para reconstruir o sinal original.

6.8.5. Parâmetros de qualidade de serviço

A transmissão de voz por pacotes em tempo-real deve satisfazer alguns requisitos, tais como garantir um atraso máximo para cada pacote, uma variação máxima do atraso dos pacotes e uma taxa máxima de perda de pacotes. Dentre estes parâmetros, o atraso é o que mais contribui para a perda da interatividade da conversação. Na tabela 6-8.3 são apresentados alguns valores de tolerância ao atraso, recomendados pelo ITU-T. O atraso total é dado pelos atrasos de codificação e decodificação, de geração de pacotes, de propagação e de espera em filas. O atraso de espera em filas considera não apenas as filas dos roteadores, mas também as filas dos *buffers* dos receptores. Na Internet, o atraso devido à espera em filas representa a maior parte do atraso total quando a rede está congestionada.

Atraso da voz	Tolerância
Até 150ms	Aceitável com boa interatividade.
150ms - 400ms	Aceitável, mas o usuário já percebe alguma perda de interatividade.
Acima de 400ms	Inaceitável, com perda de interatividade.

Tabela 6-8.3: Tolerância ao atraso

Ao contrário do tráfego de dados, o tráfego de voz admite a ocorrência de perdas de pacotes. Contudo, existe um certo limite para a taxa de perda de modo a não prejudicar a inteligibilidade. Outro fator relevante para a transmissão de voz é o número de pacotes consecutivos perdidos em um mesmo fluxo, isto é, a perda de pacotes em rajada. A perda em rajada é mais prejudicial do que a perda de pacotes intercalados, pois acarreta uma maior quantidade de silêncio.

6.8.6. Técnicas para reduzir o efeito da perda de pacotes

Existem algumas técnicas para suavizar a perda de pacotes. A grande vantagem destas técnicas é não acrescentar sobrecarga na rede. Dentre essas técnicas, destacam-se a substituição por silêncio, a substituição por ruído, a repetição de pacotes, a interpolação e a intercalação. Detalhes de cada uma destas técnicas serão visto a seguir.

A técnica substituição por silêncio consiste na substituição dos pacotes perdidos por silêncio. Esta técnica pode causar cortes na voz, sobretudo, quando são utilizados pacotes grandes ou quando há uma alta taxa de perda.

Já a substituição por ruído consiste na substituição dos pacotes perdidos por ruído branco. Esta técnica apresenta um melhor desempenho que a substituição por silêncio, pois o cérebro humano é capaz de reconstruir a mensagem recebida na presença de ruído de fundo, o que não é possível com o silêncio.

A técnica de repetição de pacote consiste na repetição do último pacote corretamente recebido no lugar do pacote perdido. É recomendável que o sinal repetido seja desvanecido para garantir uma melhor qualidade.

A interpolação utiliza informações dos pacotes vizinhos ao pacote perdido para reconstruí-lo. A interpolação, utilizando características da forma de onda dos sons anteriores e posteriores à perda, apresenta um melhor desempenho que as técnicas de substituição por silêncio e de repetição. Pode-se utilizar também os pacotes vizinhos para tentar prever o pacote perdido, tendo em vista que a voz apresenta uma grande auto-correlação.

Por fim, na intercalação os quadros de voz são rearrumados nos pacotes, de maneira que quadros consecutivos não sejam enviados em um mesmo pacote. Estes quadros devem ser ordenados no receptor antes de serem reproduzidos. Desta maneira, a perda de um pacote significa a perda de alguns quadros espaçados, tornando-se mais tolerável ao ouvido humano. O maior problema desta técnica é o acréscimo de atraso, pois a fonte deve gerar *n* pacotes que serão rearrumados antes da transmissão.

Também existem técnicas de recuperação de perdas ditas de correção direta ou automática (*Forward Error Correction* - FEC), as técnicas de recuperação por retransmissão (*Automatic Repeat Request* - ARQ) e as técnicas mistas (FEC+ARQ ou ARQ+FEC). As técnicas de recuperação de perdas por retransmissão e mista são baseadas na retransmissão dos pacotes perdidos, fazendo com que estas técnicas não sejam adequadas para tráfegos com restrições de tempo-real. A correção direta (FEC) acrescenta redundâncias aos pacotes de voz para viabilizar a detecção e a correção de pacotes perdidos.

A variação do atraso (*jitter*) também pode degradar a qualidade da voz transmitida, pois a reprodução de um tráfego de voz deve ser feita de forma cadenciada. O fator que mais contribui para o aumento da variação do atraso é a espera nas filas dos roteadores, pois o tamanho destas filas está diretamente relacionado com a carga da rede. Conseqüentemente, os pacotes de voz podem sofrer diferentes atrasos em um mesmo roteador. Outro fator que contribui para o aumento da variação do atraso é a possibilidade dos pacotes poderem

percorrer diferentes caminhos, ou seja, passar por diferentes roteadores, o que aumenta a probabilidade de ocorrerem diferenças de atraso. Para solucionar este problema são utilizados *buffers* no receptor para o armazenar a voz antes de sua reprodução. Dessa forma, quanto maior a variação do atraso, maior o tamanho do *buffer*. Assim, após a chegada no receptor cada pacote deve ser armazenado no *buffer* e aguardar o seu momento de reprodução, causando um aumento no atraso do pacote. Caso um pacote chegue após o momento de sua reprodução, ele é automaticamente descartado. Por este motivo, na escolha do tamanho do *buffer* existe um compromisso entre o atraso do pacote e a taxa de descarte. Um *buffer* pequeno pode diminuir o atraso do pacote, no entanto, se ele for muito pequeno em relação à média da variação do atraso, muitos pacotes serão descartados. Por outro lado, quanto maior for o tamanho do *buffer*, a fim de minimizar o descarte de pacotes, maior será o atraso do pacote. Desta maneira, é necessária a utilização de mecanismos adaptativos de armazenamento de voz, nos quais o tamanho do *buffer* varia de acordo com o atraso, a variação do atraso e a taxa de perda.

Telefonia Móvel Celular

7.1. INTRODUÇÃO À TELEFONIA MÓVEL CELULAR

7.1.1. Conceito de sistema móvel celular

Define-se um sistema de telefonia móvel como uma rede de comunicações por rádio que permite mobilidade contínua.

O sistema celular difere do móvel convencional na medida em que divide a área de cobertura em várias regiões menores, chamadas células, a fim de diminuir a potência transmitida e reutilizar bandas de freqüência em outras células (em geral não colindantes).

7.2. DESCRIÇÃO DO SISTEMA CELULAR

A divisão em células, criada pela Bell Labs em 1947, alavancou os sistemas de telefonia móvel. Se todo o tráfego ficasse concentrado em uma única estação central, o consumo de potência, tanto dessa estação central quanto dos terminais móveis, seria bem maior. Além disso, a capacidade de processamento da estação central teria que ser muito grande, de forma que se pudesse lidar com as várias ligações simultâneas.

7.2.1. Estrutura celular

Toda célula possui uma estação rádio-base (ERB). Há dois tipos básicos de células:

- Omnidirecionais: a ERB tem uma antena que transmite em todas as direções com a mesma potência, formando assim uma área de cobertura circular, cujo centro é a própria antena. Para facilitar os gráficos, essas regiões são comumente representadas por hexágonos;
- Setorizadas: a ERB tem antenas diretivas, de tal forma que cada uma cubra determinada área. Eventualmente, de acordo com a necessidade, uma das antenas pode transmitir com potência maior que as demais.

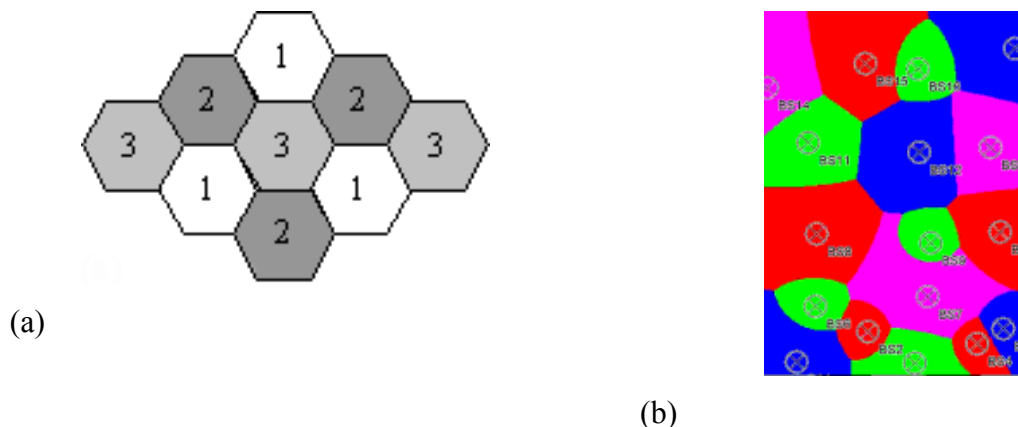



Figura 7-2.1: Estruturas das células: (a) omnidirecionais; (b) setorizadas

7.2.2. Reutilização de freqüências

Um grupo de células vizinhas forma um *cluster*, no qual cada célula tem uma parte do total de canais de voz disponíveis. Um *cluster*, em geral, não corresponde a toda a área que se deseja cobrir. Dessa forma, o que se faz é reaproveitar os canais de voz em outros *clusters*.

É necessário, todavia, calcular qual a distância mínima que se deve respeitar para poder reaproveitar os canais, de modo a evitar interferências. No caso de um sistema homogêneo, isto é, em que todas as ERB's transmitem com a mesma potência e as células são omnidirecionais, é comum que a distância mínima D seja calculada como:

$$\frac{D}{r} = \sqrt{3N}$$

Onde r é o raio da célula e N é o número de células no ***cluster*** em questão.

O número de canais em cada célula pode ser fixo (o que é ineficiente, visto que pode haver desperdício de canais em uma célula e ausência de canais em outra) ou variável. Neste caso, células podem "emprestar" canais que não estejam utilizando para outras do mesmo *cluster*. Esses "empréstimos" são controlados por uma entidade superior, chamada de MSC (*mobile switching center*). Há ainda uma terceira possibilidade, melhor que as outras duas. Nela, nenhuma célula tem canais *à priori*. Quando uma ligação é requisitada, a ERB correspondente manda um sinal para a MSC, que lhe fornece ou não um canal. De fato, o tráfego aumenta, mas o controle dos canais fica centralizado na MSC, o que torna essa estratégia de distribuição mais barata e confiável.

7.2.3. Métodos de acesso múltiplo

Existem três tipos de acesso múltiplo à rede de telefonia celular:

- TDMA (*time division multiple access*): o tempo é dividido em pequenos intervalos, chamados de *slots*. Em cada *slot*, um único usuário lança mão do canal; todos os outros se mantêm em silêncio. Essa característica é um limitador do número de usuários simultâneos dessa tecnologia. O funcionamento do TDMA se baseia na característica do ouvido humano de não perceber períodos curtos de silêncio. Se muitos acessam a rede ao mesmo tempo, pode acontecer que o intervalo entre um *slot* e outro do mesmo usuário seja grande demais, de modo que o receptor da mensagem o perceba. Além disso, é imprescindível que todos os elementos do sistema mantenham-se sincronizados;

- FDMA (*frequency division multiple access*): nessa tecnologia, cada usuário fica com um canal dentro da banda total do sistema. Esse canal fica "preso" ao usuário e só é liberado quando a ligação é terminada ou quando se faz uma troca de célula (*handoff* – ver abaixo). Esse sistema também apresenta uma limitação quanto à quantidade de acessos simultâneos: a banda do sistema é finita e, por conseguinte, também o é o número de canais. Logo, pode acontecer de todos os canais ficarem ocupados e o sistema ficar sobrecarregado. Existe também a necessidade de ortogonalidade entre as portadoras dos diversos canais, de modo a evitar interferências entre elas;

- CDMA (*code division multiple access*): a divisão entre os sinais dos vários usuários é feita por códigos. Pode ser de dois tipos:

    - FH-CDMA (*frequency hopping CDMA*): usa características do TDMA – divisão do tempo em *slots* – e do FDMA – cada usuário com um canal de freqüência. De fato, a cada *slot* a transmissão de um determinado usuário se dá por um canal diferente. A mudança entre as freqüências (*frequency hopping*) de transmissão segue um código preestabelecido entre transmissor e usuário;

- o DS-CDMA (*direct sequence CDMA*): cada usuário do sistema possui um código privado e único. Conhecem-no apenas o terminal e a central. Os sinais enviados por ou a um usuário são multiplicados por seu código, que tem uma taxa bem superior à dos sinais originais.

7.3. REDE DE TELEFONIA CELULAR

7.3.1. Partes integrantes

Existem três elementos básicos nas redes celulares em geral: o terminal móvel, a ERB e o centro de comutação e controle.

O terminal, mais comumente chamado de telefone celular, contém uma unidade de controle, um transceptor (dispositivo que funciona tanto como transmissor quanto como receptor) e uma antena. Ele transmite e recebe sinais de voz, de dados e de controle. De tempos em tempos, o terminal envia um sinal que é captado pela ERB mais próxima a ele, de modo ao sistema saber onde ele se encontra.

A ERB possui um grupo de canais, uma interface de comunicação com a central e antenas. Ela monitora os terminais dentro de sua área de cobertura e intermedia a comunicação entre a central e os terminais. A comunicação entre uma ERB e os terminais dentro de sua célula pode ser feita por um dos seguintes métodos:

- FDD (*frequency division duplex*): a transmissão no sentido ERB-terminal (chamada de *downlink*) é feita por uma faixa de freqüência diferente da transmissão no sentido oposto (conhecida como *uplink*);
- TDD (*time division duplex*): o tempo é dividido em *slots*. O *downlink* e o *uplink* se dão em slots diferentes.

A central é responsável por monitorar todo o tráfego das ERB's dentro de sua área de atuação, comutar as ligações originadas ou destinadas a um terminal dentro dessa área, comunicar-se com outras centrais de sistemas celulares ou fixos, entre outras coisas. É importante notar que, mesmo que o terminal de origem e o de destino estejam dentro de uma

mesma célula, obrigatoriamente a comutação é feita pela central. É nela, inclusive, que se encontra a MSC. A ligação entre ERB's e centrais é feita por microondas ou fibra ótica.

7.3.2. Funções características

A) *Handoff*:

Quando um terminal move-se em direção à fronteira de duas células, o sinal da ERB da célula original torna-se fraco. Antes que ele atinja o nível mínimo de qualidade aceitável, que é em torno de (–95 ± 5) dBm, e, conseqüentemente, a ligação seja interrompida, deve-se trocar a comunicação para a célula de destino. Esse processo é chamado de *handoff* e é controlado pela MSC.

É necessário calcular bem o ponto onde o *handoff* deve ser feito. Se esse ponto ficar muito antes da divisão entre as células, as trocas de ERB's podem ficar constantes, o que sobrecarregaria a MSC. Se ficar próximo demais, corre-se o risco de o sinal ficar muito fraco antes da mudança da célula e a ligação cair.

Durante uma conversação, é comum acontecerem enfraquecimentos momentâneos no sinal que chega à ERB. Tais enfraquecimentos podem ser causados por interferências ou quaisquer outros problemas no canal. Para impedir que se faça *handoff* desnecessariamente devido a um desses enfraquecimentos momentâneos, mede-se a média da potência dos sinais. Nos sistemas analógicos, há um medidor junto a cada ERB que monitora e repassa para a MSC os dados da potência dos sinais recebidos. Essa se encarrega de providenciar o *handoff*. Nos sistemas digitais, é o terminal que mede o sinal das ERB's ao seu redor e repassa a informação para a ERB com maior potência. A MSC fica liberada e o *handoff* fica mais rápido.

Nos sistemas CDMA, usa-se o *soft handoff*. Nele, o terminal, ao invés de trocar completamente de ERB, fica durante um tempo recebendo e transmitindo informações para duas ERB's. Isso faz com que a probabilidade de queda da ligação seja minorada.

B) *Roaming*:

Faz-se *roaming* toda vez que um terminal move-se para uma área coberta por uma central diferente da sua. O procedimento automático funciona assim:

1. O terminal registra-se na nova central;
2. A central verifica se o terminal já havia se registrado anteriormente. Caso contrário, ela comunica à central original sua nova posição;
3. A central de origem do terminal registra qual central que seu assinante está visitando.

As ligações destinadas àquele terminal irão para sua central nativa. Lá, verificar-se-á que o terminal não se encontra dentro daquela área, mas sim em uma outra. Nessa outra central, o terminal recebe um número fictício que o identifica. Quando o terminal deixa a área dessa central, esse número fica livre para ser utilizado por outro terminal visitante.

7.3.3. Projeto de sistemas celulares

O projeto de sistemas celulares envolve aspectos como definição do padrão de reuso, interferência co-canal, limiar da relação sinal/ruído aceitável, previsão de tráfego na Hora de Maior Movimento (HMM) por região, a distribuição geográfica dos usuários, dentre outros que podem ser analisados separadamente e depois relacionados entre si. Outros aspectos estão fora do controle do projetista e são estudados como estratégia de implementação.

Assim, um projeto pode ser dirigido pela tecnologia. Desta forma as características do sistema são ditadas pela tecnologia e equipamentos disponíveis no mercado. Isto traz a grande vantagem de um cronograma de implementação curto. Por outro lado o usuário deve adaptar suas necessidades aos equipamentos existentes. Já em uma outra abordagem, o projeto dirigido pelas aplicações busca o entendimento das necessidades do usuário. As atenções estão voltadas para o grau de serviço (GOS) esperado, a qualidade de serviço (QoS) contratada, o tipo de informação a ser trafegada, largura de faixa disponível ao usuário, a privacidade na comunicação, as características do tráfego da informação (velocidade, duração, atraso), etc. Isto tudo pode até sair barato se o usuário estiver disposto a pagar pelo serviço.

Para fins de custo procura-se projetar um sistema celular com a quantidade mínima de ERBs cobrindo a área definida, GOS e qualidade compatíveis com o anseio do usuário. Deve-se considerar que as áreas de maior tráfego devem ter maior atenção no projeto. Os parâmetros como área efetivamente coberta, GOS, grau de mobilidade e qualidade do sinal recebido são conceitos relevantes quando a satisfação do usuário é o fim.

O planejamento de um sistema começa pela definição da área de serviço a ser atendida a partir da distribuição geográfica do tráfego a ser atendido. Em seguida, em ambiente computacional de simulação, localiza-se a primeira ERB. A partir de um plano de reuso, localizam-se as outras ERBs em função do tráfego oferecido por cada uma delas. Todo sistema deve ser projetado para permitir expansões tanto em área atendida como em tráfego oferecido.

O ambiente computacional faz a predição de cobertura celular e detecta as possíveis condições de interferências co-canal. A partir deste resultado o projetista rearranja as ERB de modo a evitar as condições de interferência. Após nova predição a equipe de projeto começa os levantamentos em campo. Observando-se a coerência no projeto, inicia-se a implantação do sistema.

Os passos descritos formam uma linha geral de projeto de sistema de comunicação móvel celular.

7.4. EVOLUÇÃO DO SISTEMA CELULAR

Os primeiros sistemas de telefonia celular nos moldes que se conhecem hoje foram implantados nos Estados Unidos (AMPS) e no Japão (NTT), em 1979. Um ano depois, os países nórdicos lançaram seu sistema, o NMT, e após dois anos, foi a vez do Reino Unido implantar o TACS. No Brasil, o primeiro sistema foi o AMPS, introduzido pela Telerj Celular, em 1990. Esses primeiros sistemas ficaram sendo conhecidos como a primeira geração da telefonia celular. Eram todos analógicos: usavam FM para modular a voz.

A segunda geração de celulares nasceu no início dos anos 90. Trouxeram consigo os sistemas digitais. As principais vantagens desses em relação aos analógicos são:

- Melhoria da eficiência espectral: permitiu que o acesso seja feito por multiplexação do tempo ou de códigos. Dessa forma, mais de um usuário pode utilizar mesma faixa de freqüência;
- Rejeição à interferência: conseqüência das técnicas de modulação digital;
- Utilização de códigos corretores de erro: para a mesma potência das ERB's, melhor qualidade de serviço;
- Barateamento do *hardware*, o que possibilitou o grande crescimento da telefonia celular em todo o mundo.

Em 1997, foram implantados no Brasil os primeiros sistemas de telefonia celular digital. No ano seguinte, o governo federal privatizou as empresas de telecomunicações. Depois disso, a telefonia móvel explodiu e hoje (2003) existem mais celulares que telefones fixos no país.

7.5. PADRÕES DE TELEFONIA MÓVEL CELULAR

Nessa seção, serão caracterizados os principais padrões de telefonia móvel celular encontrados no Brasil nos primeiros anos do século XXI.

7.5.1. AMPS

O AMPS (*Advanced Mobile Phone Service*) foi desenvolvido pela Illinois Bell, empresa de telecomunicações daquele estado americano, em cuja capital, Chicago, o primeiro padrão de telefonia celular foi implantado comercialmente.

A tabela abaixo apresenta algumas características do padrão:

Bandas de FDD	*Downlink*	870 a 890 MHz
	Uplink	825 a 845 MHz
Espaçamento entre canais		30 kHz
Número de canais		666
Raio de cobertura		2 a 20 km
Modulação do sinal de áudio		FM
Modulação do sinal de controle		FSK
Taxa de transmissão de dados		10 kbit/s

Tabela 7-5.1: Características do AMPS

No Brasil, terminais analógicos não são mais comercializados. No entanto, as operadoras são ainda obrigadas a manter suas centrais e ERB's analógicas funcionando. Há dois motivos para isso: o número significativo de terminais analógicos ainda existentes, principalmente fora dos grandes centros, e a falta de cobertura total dos padrões digitais, o que obriga a realização de *roaming* analógico (o celular digital passa a operar em AMPS).

O padrão AMPS evoluiu para o D-AMPS, também conhecido por TDMA.

7.5.2. GSM

Surgimento

A tecnologia GSM (*Global System for Mobile Communication*) foi desenvolvida com a finalidade básica de integrar em um único sistema móvel os vários países da Europa. Na época do desenvolvimento do padrão, existiam no continente diversos sistemas em operação que eram incompatíveis entre si. O novo sistema deveria permitir a livre movimentação entre os diversos países, o que ocorre freqüentemente na Europa. Além disso, havia uma demanda por serviços de comunicações que excedia a capacidade dos sistemas analógicos em uso à época.

Desde 1978, a banda de freqüências em torno de 900 MHz estava reservada para comunicações móveis na Europa. Em 1981, o celular analógico foi lançado e, quase ao mesmo tempo, houve um estudo conjunto entre França e Alemanha voltado à tecnologia celular digital e à possibilidade da criação de um sistema pan-europeu. Em 1982, um comitê de trabalho especial, o *Groupe Spécial Mobile* (GSM) foi criado para analisar e continuar o estudo franco-germânico.

O sistema a ser desenvolvido tinha alguns requisitos a atender:

- Ser compatível com ISDN;
- Ter boa qualidade de sinal de áudio;
- Usar eficientemente as freqüências de rádio;
- Ser compatível com outras redes de comunicações de dados;
- Ter alta capacidade (muitos usuários ou taxas elevadas);
- Ter bom nível de segurança para os assinantes.

Em 1986, o comitê de trabalho deu um passo à frente com o estabelecimento de um núcleo permanente de pessoas designadas para a continuação do trabalho e a criação de normas para um sistema digital do futuro. Aproximadamente um ano depois, o memorando de entendimento, ou MoU, como foi denominado, foi assinado por mais de 18 países. Esse memorando declarava que os signatários participariam do sistema GSM e o colocariam em operação até 1991. Em 1989, o GSM foi transferido para a organização ETSI (*European*

Telecommunications Standards Institute – Instituto Europeu de Normas de Telecomunicações). Uma vez sob o controle do ETSI, o sistema teve o seu nome alterado para *Global System for Mobile communications*.

Em 1990 as especificações do GSM900 (da faixa de 900 MHz) foram congeladas, dando início à adaptação do GSM para a banda de 1800 MHz. Em 91, os primeiros sistemas entraram em operação. No ano seguinte, na Finlândia, a primeira rede GSM passou a funcionar.

A alocação de freqüência para o GSM 1800 (DCS1800) estabeleceu as seguintes faixas:

- 1710 – 1785 MHz (*uplink*);
- 1805 – 1880 MHz (*downlink*).

Decidiu-se por um padrão aberto com interfaces bem definidas entre as diferentes partes do sistema, objetivando o fim dos monopólios e o conseqüente aumento do mercado. Assim, pode existir competição entre as empresas que fabricam os equipamentos e as que fornecem serviços e produtos, permitindo que o custo total da tecnologia diminuísse para o usuário final.

O GSM foi concebido para ser utilizado sem causar grandes modificações à PSTN (*Public Switch Telephone Network*). Suas principais características são:

- Maior tolerância a perturbações entre células devido à transmissão digital;
- Melhor qualidade de voz;
- Suporte à transmissão de dados;
- Cifragem da fala;
- Garantia de segurança das informações do usuário;
- *Roaming* internacional;
- Existência de novos serviços graças à compatibilidade com ISDN.

Nos primeiros anos do século XXI, o GSM já era a tecnologia mais usada por celulares de todo o mundo, contando com mais de 930 milhões de usuários, inclusive em todos os países da América Latina.

Antes de operar no Brasil, foi preciso que se definissem as faixas de freqüência para a operação do GSM. Representantes norte-americanos fizeram um forte lobby para a adoção do mesmo sistema utilizado nos EUA, ou seja, a faixa de 1,9 GHz para o GSM1800, já que a faixa de 1,8 GHz já era destinada por eles para uso militar. Com isso, os EUA seriam os responsáveis pelo fornecimento de aparelhos e equipamentos para o Brasil. Porém, o presidente do Brasil (na época, Fernando Henrique Cardoso) decidiu pela adoção do sistema europeu, com a faixa de 1,8 GHz para o GSM1800, ficando o GSM900 na faixa dos 900 MHz.

Assim, a tecnologia GSM entrou em operação no Brasil em 2002 e chegou à marca de 5 milhões de usuários em apenas dezesseis meses. É considerada uma notória história de sucesso e chamou a atenção dos principais órgãos relacionados à tecnologia GSM. Nesse período, o uso da tecnologia cresceu mais de 11% ao mês, de acordo com a Anatel, o que representa cerca de 300 mil novos consumidores por mês, parte deles migrando de sistemas CDMA e TDMA.

O desenvolvimento do GSM

Fase 1 – Serviços básicos:

- Conversação;
- Dados (comutação por circuito), com até 9,6 kbps, fax, encaminhamento e bloqueio de chamadas, colocação em espera;

Com o aumento da demanda por serviços de dados nas redes de telefonia do mundo todo, o GSM passou a desenvolver em sua plataforma os caminhos para disponibilizar dados em várias taxas para as operadoras. Esses caminhos tiveram que ser desenvolvidos de forma gradual e econômica, tal que as operadoras pudessem evoluir suas redes sem perder seus investimentos anteriores.

Fase 2 – Serviços avançados:

- *Value added service* (VAS): *short message service* (SMS) – implantado e, posteriormente, aprimorado – e correio de voz;
- Serviços suplementares adicionais;
- Recomendações congeladas em 1995;
- GSM900 e GSM1800 combinados;
- Aprimoramentos no acesso de rádio e nos cartões SIM.

As interfaces abertas – mas com regulamentações – limitam em parte a evolução, mas permitem a existência de diferentes fornecedores. São interfaces abertas no GSM:

- MS – BS (BTS) → interface aérea;
- MSC – BSC → interface A.

As demais interfaces não são completamente abertas, pois sistemas comerciais foram lançados antes das especificações.

Em 1998, houve os primeiros testes da fase 2, em Cingapura. A tecnologia GSM alcançava à época cerca de 120 milhões de usuários. No ano seguinte, já eram 250 milhões.

Fase 2+ – Transmissão de dados:

- *Inteligent network* (IN): plataforma para criação de novos serviços e modificação dos já existentes;
- Serviços pré-pagos;
- HSCSD (*High Speed Circuit Switched Data*):
    - Usa comutação por circuito;
    - Transmite dados com até 60 kbps;
    - Ocupa boa parte dos recursos da interface aérea;
    - Usuário utiliza-se de 1 a 8 slots de tempo simultaneamente;
- GPRS (*general packet radio service*);
- EDGE (*enhanced data rates for GSM environment*).

As características da fase 2+ foram fixadas em intervalos regulares (*releases*). Os comitês subtécnicos da ETSI conhecidos por SMG1 até 12 são os responsáveis pela evolução técnica das especificações.

Além da fase 2+: UMTS (3G), 4G, 5G...

Subsistemas do GSM

O sistema GSM interage com três agentes: o operador do sistema, os usuários e as redes externas.

Para o usuário, a fronteira com o sistema é a interface de rádio, que ele acessa através de seu aparelho. Entre o usuário e o sistema existem 2 sub-sistemas: o BSS e o MS. Para as outras redes de telecomunicações, a fronteira é o MSC, que permite tráfego entre partes do sistema GSM sem comunicação direta entre si. Entre essas redes e o sistema existe outro subsistema, o NSS. Para o operador, essa fronteira está entre as estações de trabalho e os funcionários responsáveis pela operação e manutenção do sistema. Entre o operador e o sistema existe um subsistema: o OSS.

- MS (*mobile station*): estação móvel. Único equipamento percebido pelo usuário. Responsável pela transmissão e recepção de sinais de áudio, faz a interface com o usuário através de microfone, alto-falante, *display* e teclado. Faz também interface com outros equipamentos, como fax e computador. É composto por duas partes:

  - ME (*mobile equipment*): o aparelho, a parte que faz a interface com o usuário;
  - SIM (*subscribe identity module*): cartão que contém todas as informações relacionadas à assinatura (no lado do usuário). Responsável por informações de segurança. A interface entre SIM e ME é plenamente especificada.

- BSS (*base station sub-system*): estação rádio-base (ERB). É controlada por um MSC, que por sua vez controla vários BSS's. Reúne os equipamentos relacionados à rádio-freqüência. Faz o contato com a MS através da interface aérea. Faz também o contato com os comutadores do NSS. Tem a função final de conectar a MS com o NSS. Além

disso, gera dados estatísticos que são enviados ao OSS para pós-processamento. É dividido em:

- BTS (*base transceiver station*): reúne os transceptores e os rádios. Faz também o controle da potência de transmissão, a organização do *paging* (sinal transmitido a todas as áreas de localização e que contém a identificação do assinante, sendo que todas as MS da área recebem o sinal, mas só a MS objetivada responde, estabelecendo a ligação ponto a ponto) e o controle de *handoff*. Os transceptores são responsáveis pela codificação de canal, pela cifragem, pela modulação dos dados e pelas medições de potência dos sinais e da interferência. A rede GSM permite até 16 transceptores por BTS, embora o comum seja utilizar até 4 por BTS. As demais partes da BTS são controladas pela O&M (*operations and maintenance*), que também faz a interface na interação com o operador do sistema. A comunicação com a BSC é feita por dois canais PCM de 2 Mbits/s;

- BSC (*base control station*): faz a comutação dos canais de tráfego (interface A) com os canais corretos na interface A-bis (interface entre a BSC e as BTS's). Tem outras funções, como o gerenciamento remoto da interface de rádio, a alocação e a liberação de canais de rádio e o gerenciamento de *handoff*. É conectado às diversas BTS's e ao NSS. Gerencia as BTS's, separando-as do resto da rede, e os TRAU's. A MSC sincroniza o BSC, que sincroniza as BTS's associadas ao BSC. Essa sincronização é fundamental para a manutenção de uma conexão, uma vez que se utiliza da tecnologia TDMA;

- TRAU (*transcoder/rate adapter unit*): realiza a codificação e a decodificação dos sinais de fala. Normalmente, localiza-se próximo ao BSC ou MSC.

- NSS (*network and switching sub-system*): faz as principais funções de comutação. Contém os bancos de dados necessários para o gerenciamento de mobilidade e controle das assinaturas. Tem como função gerenciar a comunicação entre os usuários GSM e usuários de outras redes. Divide-se em:

- MSC (*mobile switching center*): central de comutação. Existe também o G-MSC (gateway MSC), que faz a interface com outras redes (PSTN, ISDN);

- HLR (*home location register*): banco de dados que contém as informações referentes às assinaturas residentes à área. Contém o AuC (*authentication centre*), que gerencia os dados relacionados com a segurança dos assinantes;

- VLR (*visitor location register*): banco de dados que contém informações relativas à localização dos assinantes visitantes à região e informações para realizar a autenticação e cifragem. É normalmente associado a um único MSC;

- EIR (*equipment identity register*): gerencia os ME's, podendo detectar ME's que não estejam funcionando adequadamente, impedir seu acesso ao sistema e bloquear o acesso a equipamentos roubados (fica claro que é necessária uma cooperação entre as operadoras para que as informações sobre equipamentos a serem barrados sejam compartilhadas). Ele contém um banco de dados com uma lista branca, que identifica equipamentos aprovados, uma negra, para terminais roubados ou com problemas técnicos, e uma cinza, para equipamentos em análise. É opcional, isto é, a operadora pode decidir se o instala ou não.

A interconexão entre esses subsistemas é baseada no protocolo SS7.

- OSS (*operating sub-system*): responsável pela manutenção e pela operação da rede. Gerencia as configurações e o desempenho da rede, bem como os defeitos e as assinaturas. É ele que faz as cobranças. Gerencia também as estações móveis através do EIR.

Características gerais do sistema GSM

O sistema GSM utiliza um híbrido entre TDMA-FDMA: FDMA entre células e TDMA dentro delas. Utiliza canais com largura de banda de 200 kHz, que são compartilhados

através de TDMA entre 8 usuários. Utiliza modulação GMSK, com uma taxa de bits de 270 kbits/s, e FDD para a comunicação entre a ERB e um terminal.

Na codificação de canal, é utilizado um código convolucional, para a correção de erros, um *fire code*, para a detecção e correção de erros em seqüência, e um código de paridade, para detecção de erros em geral. Diferentes canais são utilizados para transmitir diferentes fluxos de dados (tráfego, aviso de chamada para um assinante, mensagens para toda a célula, etc).

A localização de usuários na rede é feita utilizando as áreas de localização (LAI – *location area identity*), onde o assinante indica cada mudança de LAI para a rede GSM. A atualização da localização do usuário é feita ao se ligar o terminal, de tempos em tempos e ao se entrar em uma nova área de localização. Normalmente, lugares como aeroportos são escolhidos para tentativas de clonagens de celulares por serem locais onde o usuário liga a MS, que estava desligada durante o vôo, e, por mudar de localização, há a necessidade de troca de informações. O GSM possui um mecanismo de defesa que é explicado abaixo.

O mecanismo de autenticação utilizado é a senha (PIN – *personal identity number*), que é conhecida apenas localmente no terminal, isto é, não é transmitida em momento algum para nenhum outro elemento da rede. O PIN é verificado pelo SIM, que pode ser bloqueado após três tentativas frustradas de digitar a senha correta. O desbloqueio só pode ser feito pela operadora. A autenticação é realizada pela verificação do SIM, algo como realizar uma pergunta que só o SIM correto pode responder. Isso é feito através de uma chave secreta, um algoritmo de identificação, um resultado do algoritmo e um número de 128 bits gerado aleatoriamente. A rede pergunta para o SIM: “Com esse número gerado aleatoriamente, qual o resultado correto do algoritmo?”. Por esse motivo, diz-se que os celulares GSM não podem ser clonados. Por outro lado, em caso de roubo ou perda, apenas o SIM pode ser bloqueado, não o aparelho.

A cifragem consiste na operação de XOR entre os bits a serem transmitidos e a seqüência de cifragem, que, por sua vez, depende de um algoritmo conhecido pelo terminal e pela ERB.

Para proteger a identidade do assinante, uma identificação temporária é utilizada, a TMSI (*temporary mobile subscriber identity*), que é alocada quando a MS se registra pela primeira vez em uma LAI e é descartada quando o usuário sai da LAI. A TMSI é armazenada na VLR que está no momento servindo ao usuário e não no HLR.

Os algoritmos utilizados, tanto para a autenticação quanto para a cifragem, variam de uma operadora para outra.

GPRS

A grande demanda por serviços além dos de voz fez com que surgisse uma geração intermediária entre a segunda e a terceira: a 2,5 G. É nessa "geração" que se encontra o GPRS.

Ao contrário do HSCSD, tecnologia de transmissão de dados utilizada no GSM inicialmente, o GPRS utiliza comutação por pacotes, em que não há conexão permanente. Assim, o canal é utilizado apenas quando há dados a serem transmitidos e há uma melhor utilização dos recursos da interface aérea, que é compartilhada com o GSM. Possui integração com redes IP e X.25, com acesso à internet e à intranet.

A introdução do GPRS praticamente não alterou os blocos principais do GSM já apresentados. Na realidade, incluiu-se uma placa na BSC, denominada PCU (*packet control unit*), que disponibiliza o tráfego por pacotes através dos novos servidores na rede denominados SGSN (*serving GPRS support node*) e GGSN (*gateway GPRS support node*) e atualizações de software para CCU (*channel codec unit*) nas BTS's.

Com o GPRS, consegue-se atingir taxas de transmissão de dados de até 150 kbps, consideravelmente melhores que no GSM, e o usuário fica sempre conectado. Como o GPRS comunica-se por pacote, pode-se basear a cobrança ao usuário na quantidade de dados transmitidos e não no tempo de conexão. Utiliza entre 1 e 8 *slots* de tempo e tem 4 modos de operação: 9,04 kbps, 13,4 kbps, 15,6 kbps e 21,4 Kbps por *slot*.

A rede GPRS consiste de 4 blocos: BSS (comum à rede GSM), SGSN, GGSN e redes externas.

O SGSN possui o mesmo nível hierárquico que o MSC na rede GSM. Ele mantém o registro da localização atual do assinante, contém funções de segurança e de controle de acesso. É conectado ao BSS por *frame relay*.

O GGSN realiza a conexão com redes externas de comutação de pacotes (por exemplo, a internet) e comunica-se com a SGSN através de IP e UDP.

A rede GPRS tem informações contidas nas HLR's e pode ter também nas VLR's.

EDGE

O EDGE também se insere entre as respostas dos desenvolvedores de tecnologia ante os anseios por melhores serviços de dados. Sua taxa ainda é aquém da especificada para a terceira geração, mas o EDGE é, certamente, seu estágio imediatamente anterior, pelo menos ao que tange à evolução do GSM.

Essa tecnologia inclui tanto comutação por circuitos (ECSD – *enhanced circuit switched data*), com três novos canais de tráfegos (E-TCH – *enhanced traffic channels*), como comutação por pacotes (EGPRS – *enhanced general packet radio service*). Utiliza GMSK e 8-PSK e codificação na interface aérea (quatro esquemas para GMSK e quatro para 8-PSK). Usa também tráfego assimétrico, ou seja, TDD com diferente número de *slots* de tempo no *downlink* e *uplink*. O acesso à rede é feito por TDMA, com 8 *slots* por quadro, o que o faz compatível com GSM. Pode chegar a taxas de transmissão de dados de até 384 kbps. Para aumentar a confiabilidade, utiliza células menores que o GSM convencional.

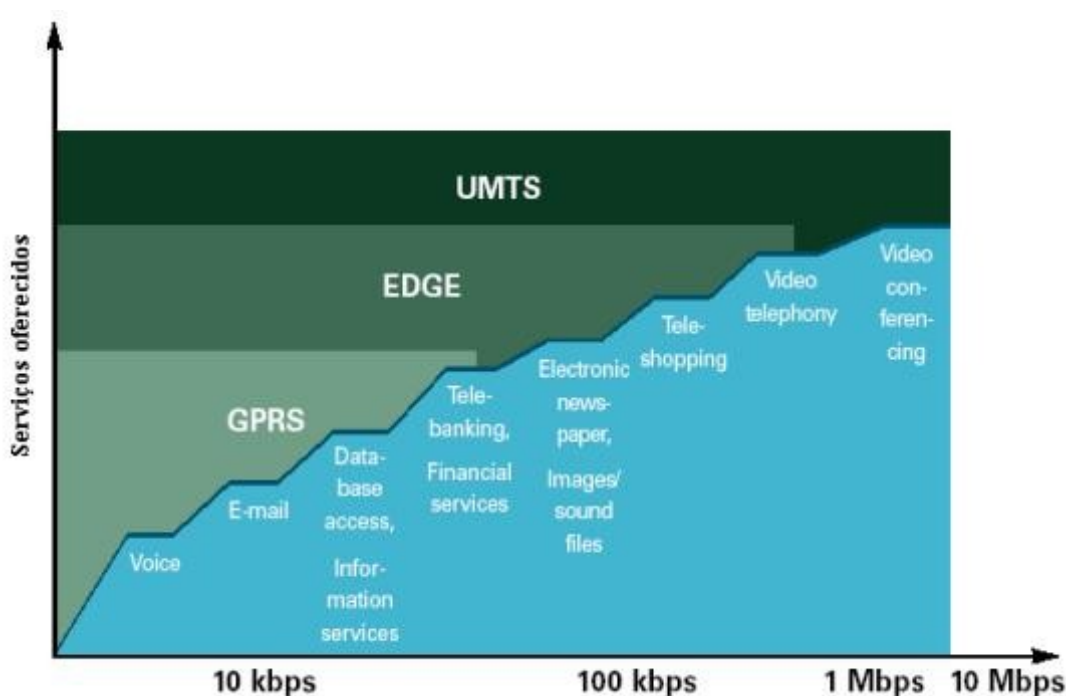



Figura 7-5.1: Serviços disponibilizados nas redes celulares de acordo com a tecnologia empregada

Modulação GMSK [traduzido (e melhorado) do texto da Universidade de Hull disponível na página de telefonia. Informações mais completas no livro do Haykin]

Para tentar entender o que é uma modulação GMSK, vamos começar lembrando o que é uma modulação em fase (PSK, também chamada, enquanto se tratar de apenas dois níveis de informação, BPSK).

BPSK (Binary Phase Shift Keying)

A representação matemática de um BPSK pode ser:

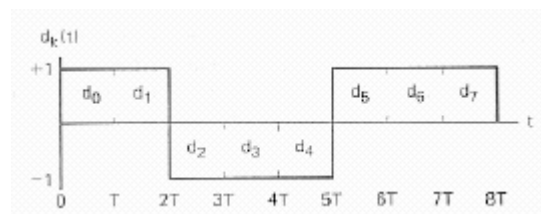
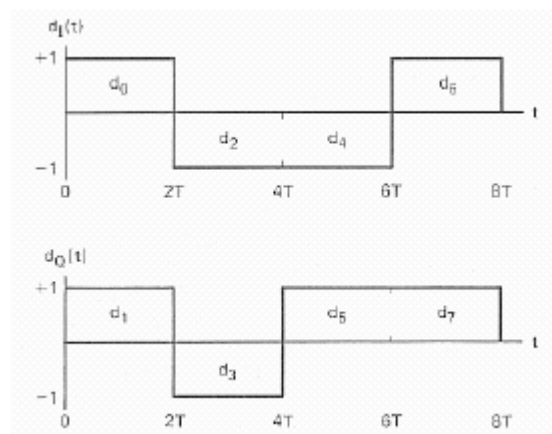
$S_o(t) = A \cos(\omega t)$ representando o nível 0
e
$S_1(t) = A \cos(\omega t + \pi)$ representando o nível 1

Onde é fácil de notar que haverá uma transição de 180º toda vez que se passar da transmissão de um "0" para a de um "1".

Uma forma de se transmitirem mais de dois bits de informação a cada transição é o uso de um maior número de fases...Exemplificando com um QPSK:

QPSK (Quadrature Phase Shift Keying)

Num QPSK, conforme os bits vão chegando ao modulador (a uma taxa 1/T bits/seg), eles são separados em duas seqüências chamadas "in phase" e "quadrature":

$d_I(t) = d_0, d_2, d_4,...$
e
$d_Q(t) = d_1, d_3, d_5,...$

representação de uma seqüência aleatória de bits

representação dos bits das seqüências par, $d_I(t)$ e da seqüência ímpar, $d_Q(t)$

Uma forma ortogonal de representar matematicamente um QPSK pode ser:

$$s(t)=1/\sqrt{2}\ d_I(t)\cos(2\pi ft+\pi/4)+1/\sqrt{2}\ d_Q(t)\sin(2\pi ft+\pi/4)$$

Que, usando identidades trigonométricas, pode ser escrita como:

$$s(t)=A\cos[2\pi ft+\pi/4+\theta(t)].$$

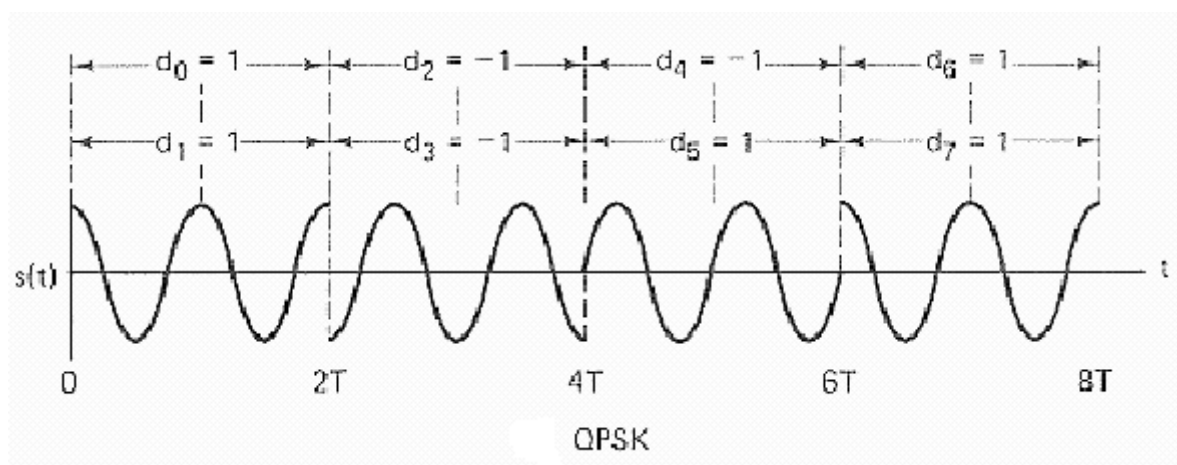
A seqüência de pulsos $d_I(t)$ modula da função co-seno com amplitude ± 1. Isto é equivalente a uma alteração na fase do co-seno de 0 ou π ; conseqüentemente isto forma um sinal BPSK. dQ(t) multiplicada pelo seno forma uma função ortogonal (nome que se dá a uma função senoidal deslocada de 90º de outra porque a integral do produto das duas dá zero) à anterior. A soma de ambas dá o sinal QPSK.

Cada um dos quatro valores possíveis para θ (t) = 0, π /2, π , -π /2 carrega a informação de dois bits. Logo estamos trabalhando com símbolos de dois bits cada.

Se mantivermos a mesma largura inicial para cada símbolo, a velocidade de transmissão de bits dobra sem aumento da largura de faixa ocupada pelo sinal. Isto é possível porque os sinais I e Q são ortogonais entre si e podem ser transmitidos sem interferência de um no outro.

Em forma de tabela podem resumir-se as fases do sinal transmitido como segue:

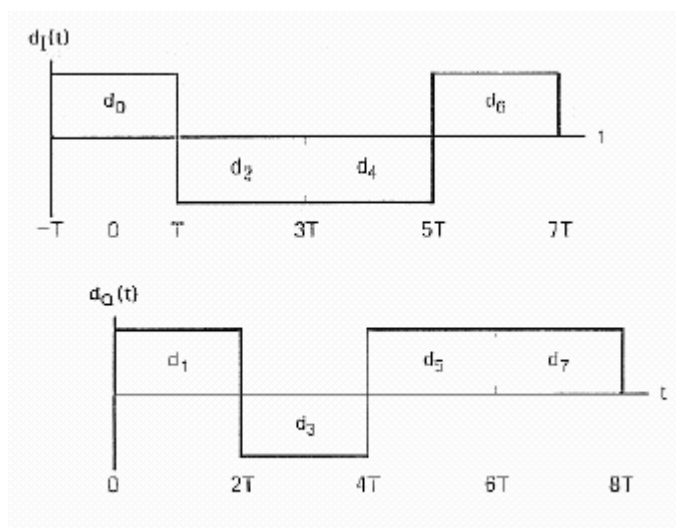
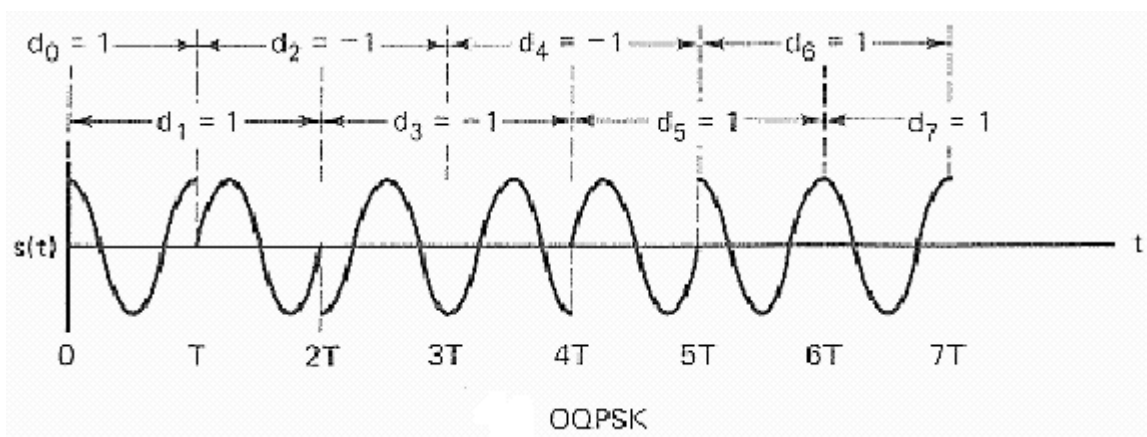
Par de bits transmitidos	Fase do sinal
1 ,1	0º
-1,-1	180º
-1,1	90º
1,-1	**-90º**

Onde se vê que em algumas transições a mudança de fase será de 180º (ao passar de 1,1 para –1,-1, por exemplo). Isto pode ser visto na figura abaixo. Estas mudanças bruscas de fase provocarão um aumento na largura de faixa ocupada pelo sinal.

Se se passa um sinal QPSK por um filtro, com a intenção de reduzir a faixa ocupada por ele, o sinal resultante já não terá uma amplitude constante e uma mudança de fase de 180^0 fará a sua amplitude cair a zero momentaneamente.

Uma maneira de evitar este efeito indesejável pode ser vista no item seguinte.

OQPSK Offset Quadrature Phase Shift Keyng

Se as duas seqüências de bits I e Q forem deslocadas de ½ intervalo de bit as flutuações de amplitude serão minimizadas uma vez que a fase nunca mudará de 180º.

Este tipo de modulação é obtido de um QPSK atrasando a seqüência ímpar de meio intervalo de bit com respeito à seqüência par, como se pode ver na figura abaixo.

A faixa de transições de fase passa a ser de 0º a 90º (a possibilidade de transição de 180º é eliminada) e ocorre com o dobro da freqüência mas com a metade da intensidade de um QPSK. As variações de amplitude continuam ocorrendo no transmissor e no receptor mas elas têm amplitude menor. A taxa de erro de bit é a mesma num BPSK, num QPSK ou num OQPSK. A figura abaixo permite ver o sinal OQPSK no tempo.

Quando se limita um sinal QPSK em banda, a interferência entre símbolos resultante obriga o envelope a diminuir um pouco na região de transição de $\pm 90^{o}$ de fase. Porém, uma vez que as transições de 180º foram evitadas no OQPSK, o envelope nunca chegará a zero como num QPSK.

O sucesso obtido com o OQPSK faz pensar na possibilidade de se diminuir ainda mais a variação na amplitude do envelope do sinal transmitido. Um jeito de se conseguir isto pode ser visto no próximo item

MSK Minimum Shift Keying

Mostrou-se antes que o OQPSK pode ser obtido pelo atraso da seqüência Q de meio bit ou T segundos com respeito à seqüência I. Este atraso não tem efeito na taxa de bits errados ou na largura de banda ocupada pelo sinal.

O Minimum Shift Keying é derivado do QPSK substituindo o pulso retangular por um pulso de meio ciclo de uma senóide. A função do tempo resultante é:

$$S(t) = d(t) \cos(\pi t/2T) \cos 2\pi ft + d(t) \sin(\pi t/2T) \sin 2\pi ft.$$

Com isto se obteria uma variação linear da fase dentro do intervalo de um bit eliminando totalmente as transições bruscas na fase e conseqüentemente diminuindo a largura de banda necessária na transmissão.

A idéia de mudar a forma do pulso pode ser aproveitada com outras ...

GMSK Gaussian Minimum Shift Keying

Um filtro com resposta ao impulso de forma gaussiana gera um sinal com baixos lobos laterais e lobo principal mais estreito que um pulso retangular. Isto pode ser aproveitado para gerar outro sinal de chaveamento de fase mínimo (GMSK).

O filtro utilizado nesta modulação terá uma largura de faixa B e estará processando um sinal cujo período do bit é T. A relação entre ambos define a largura de banda do sistema. Para o GSM estabeleceu-se que a relação usada seria BT = 0,3 com uma taxa de bit de 270,8 Kbps. Isto é um compromisso entre a taxa de erro de bits e a interferência inter-símbolos uma vez que filtros mais estreitos aumentam a interferência e reduzem a potência do sinal.

7.5.3. CDMA

Características gerais

Como foi dito na seção 7.2.3, o CDMA é um método de acesso múltiplo à rede que usa a divisão por códigos dos usuários. Existem dois tipos básicos: o FH-CDMA, no qual a freqüência de transmissão varia de acordo com um código preestabelecido, e o DS-CDMA, em que os sinais a serem transmitidos são multiplicados por um código individual, com taxa mais alta.

O FH-CDMA depende de que tanto o transmissor quanto o receptor mudem a faixa de freqüência ao mesmo tempo. Por conta disso, necessita de perfeito sincronismo. Já o DS-CDMA precisa garantir que o receptor irá decodificar exatamente o sinal que lhe foi enviado. Para isso, é necessário que a correlação cruzada entre os códigos seja pequena ou nula. O sincronismo melhora a performance, mas como é difícil controlar os atrasos, usam-se mais os sistemas assíncronos.

O DS-CDMA apresenta algumas vantagens sobre os demais métodos de acesso:

- Promove o espalhamento espectral, devido à multiplicação por uma seqüência de maior taxa, o que o faz menos sensível à seletividade do canal do que, por exemplo, o FDMA;
- Tem tempo de símbolo maior que no TDMA, o que o faz menos afetado por ecos do sinal e, portanto, passível de ser implementado com equalizadores mais simples;
- Não tem, *a priori*, qualquer limitação do número de usuários simultaneamente conectados. Sua única restrição é conseguir um número de códigos que tenham baixa correlação cruzada;
- Não necessita de reorganização de freqüências, uma vez que todas as células usam a mesma banda;
- Aproveita os ecos na recepção (receptor RAKE) para melhorar a eficiência;
- Permite o *soft handoff*, dificultando assim a perda da comunicação.

Os principais problemas apontados pelos críticos do sistema são:

- Como todos os usuários usam a mesma faixa de freqüência, eventualmente o sinal de um usuário próximo da ERB pode mascarar o sinal de outro mais distante. Por isso, o controle de potência dos terminais deve ser muito bem feito;

- Nos sistemas síncronos, a referência é dada por GPS (*global positioning system*), que é de propriedade do governo americano. Em caso de ameaça ou guerra, os EUA podem desligar seu sistema, o que faria toda a rede síncrona parar de funcionar.

Geração de sinal CDMA

O sinal do sistema CDMA é o resultado da multiplicação do *stream* de bits com o código próprio de cada usuário (figura abaixo). Porém, para que a recepção ocorra sem falhas, é necessário que os códigos sejam facilmente distinguíveis entre si. Estes códigos são conhecidos como códigos de espalhamento por serem eles os causadores do espalhamento do espectro do sinal original por toda a banda do sistema.

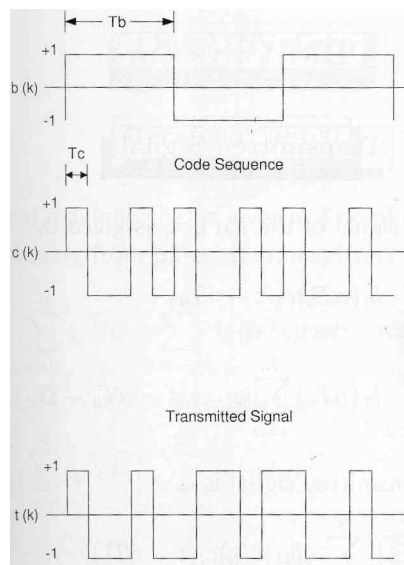



Figura 7-5.2: Geração do sinal CDMA

Dentre os códigos de espalhamento mais utilizados, estão os de Gold, de Kasami (longo e curto) e de Walsh-Hadamard.

Vale ressaltar o fato de que, para códigos ortogonais (Walsh-Hadamard), o sincronismo do sistema deve estar bem calibrado. Já para sistemas assíncronos, é necessário que os códigos de espalhamento possuam uma baixa correlação cruzada (Gold e Kasami).

IS-95

A tecnologia do espalhamento espectral, que é a base da tecnologia CDMA, era usada nas aplicações militares para, por exemplo, vencer os efeitos da interferência nas comunicações (*jamming*) e esconder o sinal transmitido. A modulação por espalhamento espectral era usada para cálculos de posição e velocidade.

A primeira empresa a ver as possibilidades dessa tecnologia para a telefonia celular foi a Qualcomm. Em 1990, ela propôs o primeiro sistema celular baseado em espalhamento espectral. Por conta disso, diz-se que a Qualcomm é "dona" do CDMA comercial.

O padrão IS-95 foi aprovado em julho de 93. Ele usa um sistema híbrido entre o DS-CDMA e o FDMA. A seqüência usada para o espalhamento espectral depende do sentido da comunicação: no *downlink*, é feito por Walsh-Hadamard; no *uplink*, usa-se Gold ou Kasami longo.

A tabela a seguir mostra um quadro comparativo entre os padrões GSM e IS-95:

Parâmetro		GSM	IS-95
Bandas de FDD	*Downlink*	824 – 849 (1805 – 1880) MHz	869 – 894 MHz
	Uplink	869 – 894 (1710 – 1785) MHz	824 – 849 MHz
Método de acesso		TDMA	DS-CDMA/FDMA
Taxa do canal codificado		22,8 kbps	19,2/28,8 kbps
Modulação		GMSK	QPSK/DPSK
Taxa de transmissão		270,8 kbps	1228,0 kbps
Espaçamento entre portadoras		200 kHz	1250 kHz

Tabela 7.5-2: Quadro comparativo entre GSM e IS-95

Evolução do IS-95

Assim como no GSM, o IS-95, também chamado de CDMA One, foi obrigado a evoluir para atender aos crescentes requisitos do mercado por taxas mais altas e melhores serviços. No padrão europeu, foi desenvolvido o GPRS e, mais tarde, o EDGE. No americano, a evolução foi o CDMA2000. Esse padrão tem como requisitos prover serviços compatíveis com os sistemas de terceira geração e também com os de segunda.

Na caminhada rumo à 3G, o primeiro passo foi o CDMA2000 1X. Esse sistema incrementa as taxas de transmissão de dados via pacotes e aumenta a velocidade da rede, bem como eleva a capacidade de tráfego de voz em aproximadamente 2 vezes, se comparado às redes CMDA tradicionais. O “1X” significa uma vez 1,25 MHz, a largura de banda padrão de uma operadora CDMA IS-95. Ao manter uma largura da banda padronizada para CDMA, permite-se que ambas infra-estruturas e, ainda mais importante, que os terminais, sejam compatíveis com ambos os padrões. Uma rede CDMA 2000 é composta de componentes de

interface aérea 1X e de um *backbone* de dados (PCN – *packet core network*). O CDMA 2000 PCN é equivalente em funcionalidade a uma rede GRPS, usando, porém, protocolos baseados em IP Móvel.

O passo seguinte é o chamado CDMA 2000 1XEV. Sua implementação será dividida em duas fases: 1XVE-DO (*data only*), voltado para tráfego veloz de dados, e 1XVE-DV (*data and voice*), que combina voz e dados em alta velocidade numa mesma portadora. A tecnologia 1XVE-DO permite velocidades superiores a 2 Mbps no tráfego de dados para usuários CDMA de uma operadora dedicada. Já a 1XEV-DV oferece alta velocidade para transmissão simultânea de dados e voz, além da possibilidade de oferecer serviços de dados em tempo real.

7.6. OFDM

7.6.1. Introdução

O sistema celular sofre do mesmo problema que todos os demais tipos de transmissão terrestre: o multipercurso, isso é, a chegada ao receptor tanto dos sinais vindos diretamente do transmissor quanto dos reflexos desses em prédios, montanhas etc. Nas ERB's, o problema pode ser satisfatoriamente resolvido com o uso de equalizadores complexos e caros. Nos terminais, no entanto, tal solução não pode ser utilizada, sob pena de inviabilizar a comercialização dos aparelhos.

Para tentar resolver esse problema, foi criada uma técnica de modulação chamada OFDM (*orthogonal frequency division multiplex*).

7.6.2. Modelagem do Multipercurso

Considere um determinado sinal s(t) representando um símbolo a ser transmitido durante um intervalo de tempo T. Num ambiente com problema de multipercurso, o sinal recebido no receptor será da forma:

$$s_{rec}(t) = A_0\, s(t) + A_1\, s(t-t_1) + A_2\, s(t-t_2) + \ldots + A_n\, s(t-t_n) \qquad \text{Equação 7.1}$$

onde A_0 é a atenuação sofrida pelo sinal principal e A_1, A_2,..., A_n são as atenuações sofridas pelas versões refletidas, e os tempos t_1, t_2,..., t_n são os atrasos sofridos por cada eco. A Figura 7.3 ilustra o problema do multipercurso.

Visto isso, o sinal que representa tal símbolo durará, no receptor, mais que o intervalo de tempo T. Na verdade, ele durará T mais o atraso sofrido pelo eco que percorrer a maior distância para chegar ao receptor. Assim, modela-se o canal por uma função de transferência g(t) de duração T_G. O resultado da ação da função de transferência g(t) sobre o sinal s(t) será um outro sinal com duração $T+T_G$.

Quando se transmite uma seqüência de símbolos de duração T por um canal modelado como o exposto acima, os símbolos posteriores serão influenciados pelos anteriores, já que tal canal "alonga" a duração de cada um deles em T_G. Isso caracteriza o que chamamos de interferência intersimbólica. Quanto maior o valor de T_G em comparação a T, maior será a interferência.

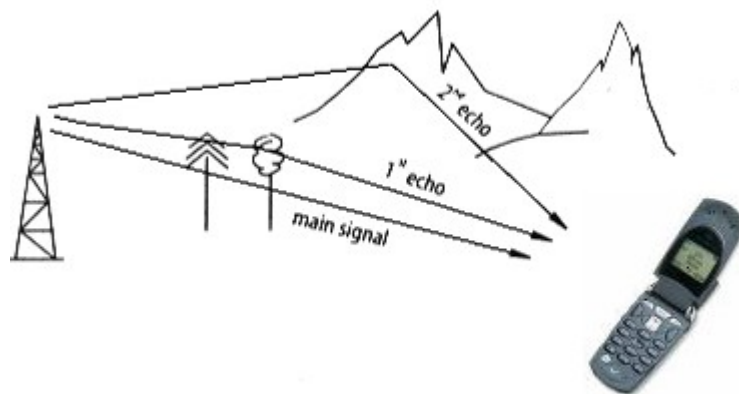



Figura 7-6.1: Problema do multipercurso

7.6.3. Multiportadoras

Para reduzir os efeitos do multipercurso, pode-se fazer com que a duração de cada símbolo a ser transmitido seja grande em comparação ao acréscimo T_G inserido pelo canal. Para que a duração T seja grande, devem-se utilizar portadoras diferentes para cada símbolo, de largura de banda razoavelmente estreita; quanto maior o valor desejado para T, menor deve ser a banda de cada portadora que carregará os símbolos.

Então, para reduzir os efeitos do canal com problema de multipercurso, deve-se usar um sistema em que cada símbolo é carregado por uma portadora de banda bem estreita. Quanto mais estreita for esta banda, maior será a duração dos símbolos e menor será a

interferência intersimbólica; além disso, mais portadoras “caberão” dentro da banda disponível a ser utilizada. Porém, não se pode reduzir infinitamente a largura de banda de cada portadora, pois nesse caso a duração dos símbolos se tornaria infinitamente grande.

Para que vários símbolos de duração T possam ser transmitidos ao mesmo tempo, cada qual carregado por sua portadora, e possam ser devidamente recuperados, essas portadoras devem ser ortogonais entre si. Duas funções são ditas ortogonais quando a integral do gráfico resultante do produto dessas duas funções (área total entre o gráfico da função e o eixo horizontal) for nula. Para exemplificar, considere as funções $c_1(t)$ e $c_2(t)$ e a função $p(t) = c_1(t)c_2(t)$, representadas na figura a seguir.

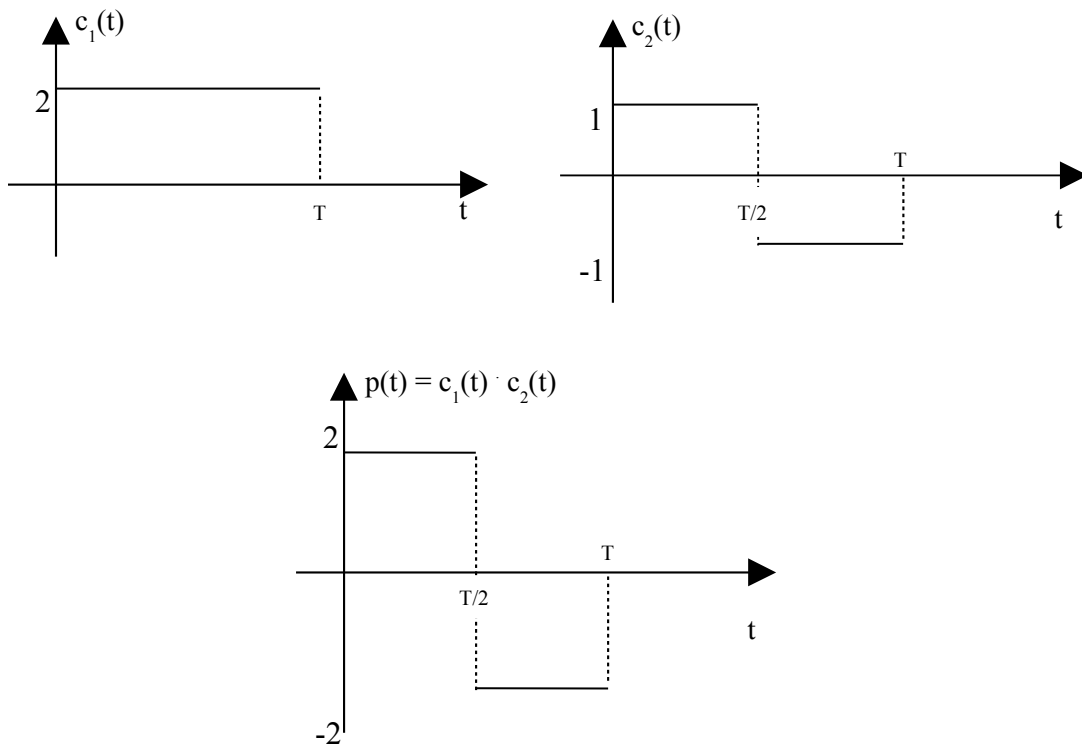



Figura 7-6.2: Funções ortogonais

No gráfico de p(t), a área acima do eixo t e a área abaixo são iguais em módulo. Como uma delas é positiva e a outra é negativa, sua soma é nula. Logo, as funções $c_1(t)$ e $c_2(t)$ são ortogonais.

Para que as portadoras que carregam os símbolos sejam ortogonais entre si, elas devem estar situadas em freqüências múltiplas de uma freqüência fundamental f_0, ou seja, as freqüências das portadoras devem se situar em $f = k \cdot f_0$, com k = 0, 1, 2, ..., n-1, onde n é o número de portadoras. A figura abaixo ilustra o diagrama de blocos do sistemas OFDM analógico.

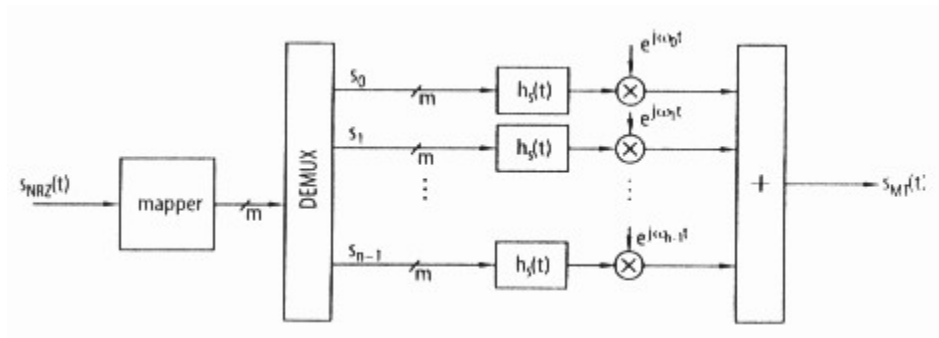



Figura 7-6.3: Diagrama de blocos do OFDM analógico

Na figura acima, $h_S(t) \cdot e^{j\theta\, t}$, onde $\theta = 2\pi\ f_0 k$ com k = 0, 1, ..., n-1, são as funções que representam as portadoras.

A figura abaixo mostra como podemos mapear uma seqüência binária, representada pelo sinal $s_{NRZ}(t)$ em portadoras OFDM. Neste caso existem dois símbolos possíveis: zero e um. A figura mostra também um símbolo OFDM no domínio do tempo – $s_{OFDM}(t)$ – formado com a seqüência de bits a ser enviada, e a representação desse símbolo no domínio da freqüência – $s_{OFDM}(f)$. Note que a portadora que carrega o terceiro bit do sinal $s_{NRZ}(t)$ não aparece nessa figura, pois o bit que a modula tem valor zero.

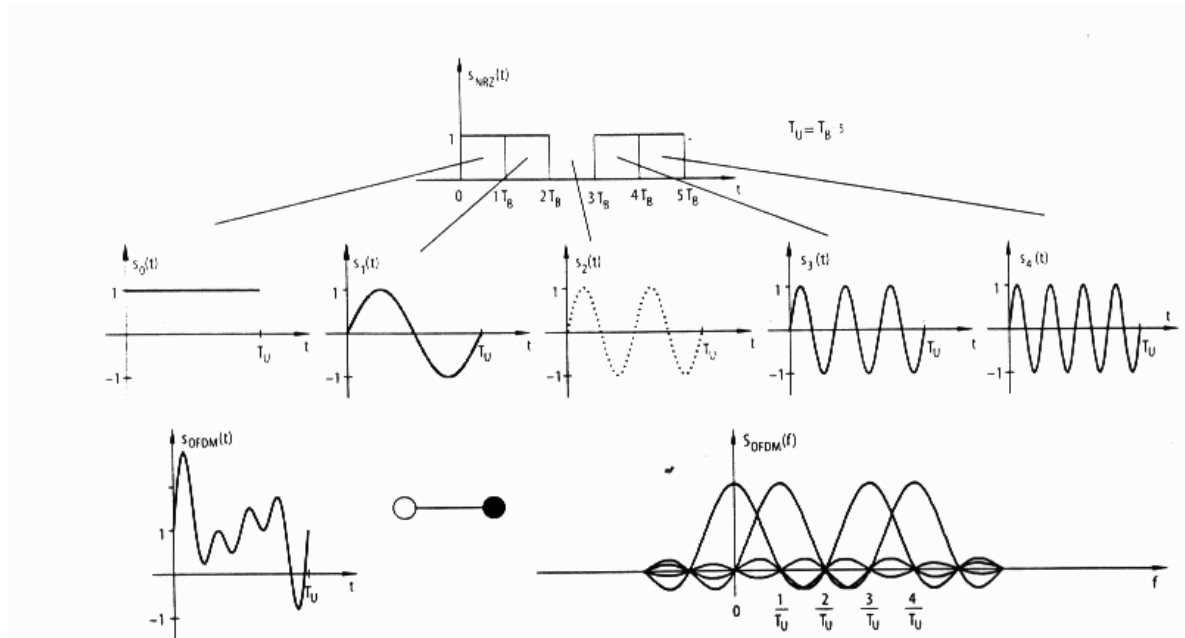



Figura 7-6.4: Representação de seqüência binária em OFDM

O fato de se transmitirem vários símbolos de duração T ao mesmo tempo, cada qual em sua portadora, não faz com que a taxa de transmissão do sistema OFDM seja maior que a de um sistema que utiliza uma única portadora com banda igual à soma das bandas das múltiplas portadoras do sistema OFDM. Isso porque, quando se aumenta a banda de uma portadora, a duração do símbolo que ela carrega diminui na mesma proporção, e vice-versa. Suponha, por exemplo, uma banda W a ser utilizada para transmissão de símbolos. Pode-se dividir essa banda entre N portadoras, resultando numa banda de W/N para cada, cada qual carregando um símbolo de duração T. Como no sistema OFDM podem-se transmitir todos esses N símbolos simultaneamente, a taxa de transmissão será N/T. Se for utilizada essa mesma banda W num sistema com uma única portadora, o único símbolo que poderá ser transmitido de cada vez terá duração T/N. Assim, esse sistema também terá taxa de transmissão N/T, como no OFDM.

Neste ponto, deve-se mencionar que a atenuação sofrida por cada portadora de um sistema OFDM é praticamente plana, visto que cada uma dessas portadoras tem largura de banda estreita. Isso torna a equalização um processo bastante fácil: caso se conheça ou se tenha como estimar o canal, a equalização é feita somente ajustando-se a amplitude e a fase de cada portadora recebida. Para fazer uma estimativa do canal, costuma-se enviar, em instantes de tempo conhecidos tanto pelo receptor quanto pelo transmissor, algumas portadoras pilotos, isto é, portadoras que possuem freqüência, amplitude e fase conhecidas. Assim sendo, o receptor compara amplitude e fase da portadora recebida com os parâmetros que ela realmente deveria ter. Desse modo, ele consegue estimar qual a distorção apresentada pelo canal para aquela freqüência determinada.

7.6.4. Prefixo Cíclico

Vimos anteriormente que não se pode reduzir infinitamente a banda de cada portadora para que a duração de cada símbolo aumente. Assim, na prática, por mais que a banda de cada portadora OFDM não seja tão grande como num sistema de portadora única, o efeito da interferência intersimbólica causada pelo canal devido ao aumento da duração do símbolo em T_G ainda existirá. Para resolver esse problema definitivamente, o símbolo enviado é modificado pela inclusão de um prefixo cíclico.

Um prefixo cíclico é uma cópia da parte final de um símbolo OFDM que é concatenada a ele em sua parte inicial. Ele também é chamado de intervalo de guarda.

Assumir-se-á que a duração deste prefixo é igual ou maior que o acréscimo no período que o canal causa no símbolo recebido, ou seja, o prefixo cíclico dura, no mínimo, T_G.

Para se entender como o prefixo cíclico pode ser usado para solucionar o problema do multipercurso, deve-se lembrar que, quando o símbolo enviado sofre a ação do canal, o resultado tem duração maior que o símbolo original. Esse aumento de duração provocaria interferência num símbolo seguinte se este fosse enviado imediatamente após o anterior. Com a inclusão do prefixo cíclico, esse acréscimo que o canal causa nos símbolos transmitidos influenciará apenas a posição ocupada por esse prefixo e o símbolo seguinte não sofrerá influência alguma. A interferência acontecerá no intervalo de tempo ocupado pelo prefixo cíclico. A figura abaixo mostra como isso acontece no domínio do tempo. O receptor recebe um sinal principal acrescido de três ecos com atrasos e atenuações diferentes. Repare que os transientes realmente acontecem no intervalo de tempo T_G. Descartando-se, no receptor, esse intervalo de tempo, podemos recuperar perfeitamente o símbolo enviado apenas ajustando sua amplitude e fase de acordo com a estimativa do canal feita com o envio das portadoras pilotos.

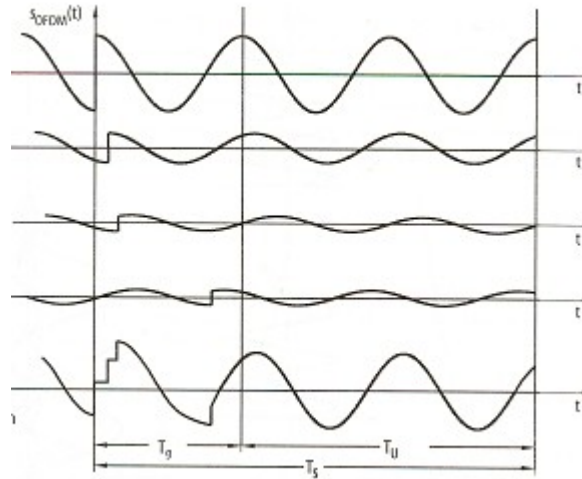



Figura 7-6.5: Sinal original e suas cópias atrasadas: o tempo T_G mostra a duração do prefixo cíclico

7.6.5. OFDM Digital

O OFDM é uma técnica de modulação digital. Sua interpretação analógica é usada apenas para facilitar o entendimento dos conceitos. Sendo assim, nessa seção far-se-á uma descrição um pouco mais formal do sistema OFDM digital.

Como foi dito anteriormente, para que as portadoras sejam ortogonais elas devem ser freqüências múltiplas de uma fundamental. Isso é equivalente, num sistema OFDM digital, a calcular uma IDFT com os valores dos símbolos a serem transmitidos. A figura abaixo ilustra o diagrama de blocos de um sistema OFDM digital.

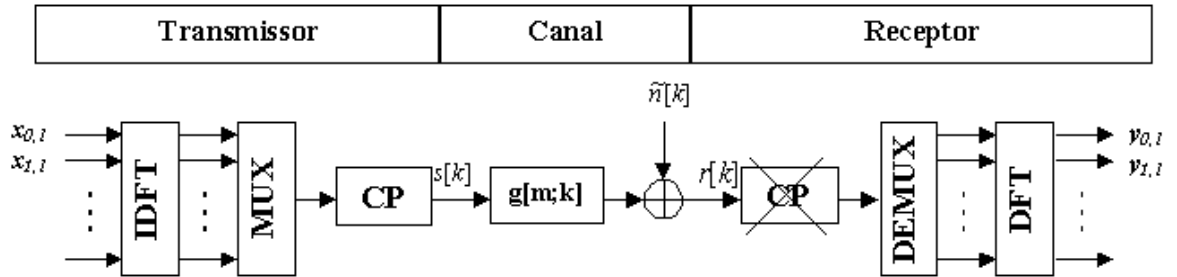



Figura 7-6.6: Sistema completo do OFDM digital

Assumindo que o sistema OFDM possua N portadoras, com uma banda total de 2π rad/s (que corresponde à banda analógica W), o sistema transmissor interpreta os símbolos (números) $x_{k,l}$ como se cada um fosse carregado por uma portadora centrada numa freqüência múltipla de 2π /N. Dessa forma, o l-ésimo símbolo OFDM (antes da inclusão do prefixo cíclico) é definido como a IDFT mostrada na figura a seguir.

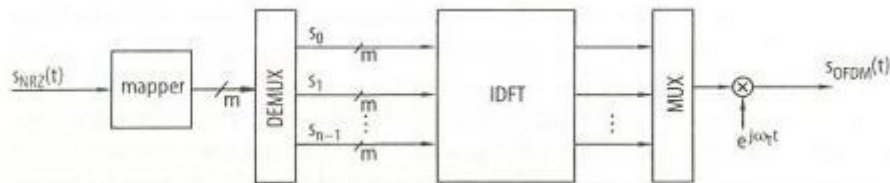



Figura 7-6.7: Esquema da transmissão do OFDM

Deve ser notado que $X_l(k)$ está espaçado no tempo de T segundos em relação a $X_{l+1}(k)$ (levando-se em conta o prefixo cíclico).

O multiplexador da figura concatena os símbolos $X_l(k)$, l = 1,2,..., inserindo entre eles o prefixo cíclico de cada um. Então, o sinal s[k] (na figura 7-6.7) terá a forma abaixo:

$$s[k] = [CP_1 \;\; X_1(0) \;\; X_1(1) \ldots \; X_1(N-1) \;\;\; CP_2 \;\; X_2(0) \;\; X_2(1) \ldots \; X_2(N-1) \;\; \ldots \; \ldots]$$ Equação 7.2

Do ponto de vista do receptor, o uso do prefixo cíclico mais longo que a resposta ao impulso g[m;k] do canal fará com que a convolução linear seja idêntica a uma convolução circular (o parâmetro m em g[m;k] representa uma possível variação da resposta do canal com

o tempo). Denotando a convolução circular por ‘⊗’, o sistema OFDM completo pode ser descrito pela seguinte equação:

$$y_l = DFT(IDFT(x_l) \otimes g_l) \quad \text{Equação 7.3}$$

onde y_l contém os N símbolos recebidos, x_l contém os N símbolos enviados e g_l é a resposta do canal ao impulso.

Usar-se-á agora o fato de que a DFT de dois sinais ciclicamente convoluídos é o produto das DFT’s dos dois sinais. Denotando o produto elemento a elemento por ‘·’, a expressão acima pode ser escrita como:

$$y_l = x_l \cdot DFT(g_l) = x_l \cdot h_l \quad \text{Equação 7.4}$$

onde $h_l = DFT(g_l)$ é a resposta em freqüência do canal.

Desde que haja uma razoável estimativa do canal, o receptor pode encontrar o símbolo emitido bastando dividir o símbolo recebido y_l pela resposta h_l do canal na freqüência em que tal símbolo está sendo carregado.

7.6.6. Aplicações práticas

Os padrões europeu e japonês de televisão digital usam o OFDM para modular os sinais na transmissão terrestre dos sinais. Não há nenhuma indicação de uso dessa técnica de modulação nos aparelhos de terceira geração. No entanto, como foi visto, é o OFDM quem melhor resolve o problema de multipercurso das transmissões terrestres. Além disso, faz parte da evolução já especificada para os dois padrões supracitados a recepção de televisão por terminais celulares. Logo, é de se esperar que, antes da entrada na quarta geração, exista algo como uma 3,5 G, que já seria compatível com recepção de TV e, portanto, usaria OFDM.

7.7. COMUNICAÇÕES PESSOAIS E SISTEMAS DE TERCEIRA GERAÇÃO

7.7.1. Introdução

A demanda por serviços de telefonia sem fio e de dados com uma cobertura satisfatória tem crescido muito nos últimos anos e a tendência é a de continuar crescendo.

Essa demanda crescente reflete-se nas atividades de engenharia, que se empenham na elaboração de novos sistemas com características mais vantajosas para o usuário, no desenvolvimento de formas de integração entre os serviços existentes e mecanismos e técnicas que permitam um aumento da eficiência na utilização dos recursos. Este desenvolvimento é necessário para atender a nova demanda não só dos usuários, mas também das novas aplicações.

Com isto, pode-se observar que as comunicações pessoais móveis representam uma descontinuidade nos serviços e tecnologias de telecomunicações, cujo desenvolvimento futuro envolverá um jogo entre as forças de mercado, as entidades regulamentadoras, as inovações tecnológicas e os padrões existentes e em estudo. O caminho para se atingir uma visão completa de serviços de comunicação pessoal será evolutivo, e a taxa e a direção desta evolução dependerão da batalha citada anteriormente.

PCS (*personal communication services*) e PCN (*personal communication networks*) são os principais serviços na área de comunicações pessoais e surgem como opção de baixo custo para os serviços celulares. O objetivo é também embutir serviços de comunicação de dados na forma de mensagens, bem como serviços de curta distância, para comunicação em ambientes fechados ou para comunicação entre prédios.

7.7.2. A rede e serviços de comunicações pessoais – PCN e PCS

O PCS (*personal communications service*) é uma designação dada para serviços pessoais de comunicações sem fio que utilizam o conceito de “a qualquer hora, em qualquer lugar”. Atualmente, esta sigla está mais voltada para o espectro de freqüências que são utilizadas pelas companhias de comunicações móveis, como a telefonia celular e o *paging*, na

posição de 2 GHz. Esta tecnologia é conhecida no Japão como PHS (*personal handy phone system*) e na Europa como PCN (*personal communication network*).

O PCS é similar ao serviço de telefone celular, mas enfatizando o serviço pessoal com grande mobilidade. Como o celular, o PCS é para usuários móveis e requer um número de antenas para atender uma área de cobertura. À medida que o usuário se move, o sinal do telefone é captado pela antena mais próxima e então repassado para uma estação base que o conecta a uma rede com fio. O aparelho para PCS é ligeiramente menor que o aparelho celular.

O "pessoal" no PCS distingue esse serviço por enfatizar que, diferentemente do celular, que foi projetado para uso nos carros, com transmissores que possibilitam cobertura de estradas e rodovias, o PCS é projetado para mobilidades maiores do usuário. Isto geralmente requer mais células transmissoras para cobrir a mesma área, mas tem a vantagem de existirem menos lugares não cobertos. Tecnicamente, sistemas celulares nos EUA operam na faixa de freqüência de 824-849MHz; o PCS opera na faixa de 1850–1990 MHz.

Diversas tecnologias são usadas para PCS nos EUA, incluindo TDMA, CDMA e GSM. Na Europa e em outros lugares, o GSM é mais utilizado.

Como se disse anteriormente, PCN é o nome que se dá a este sistema na Europa e foi introduzido na Inglaterra. O padrão tem o nome de DCS-1800 e tem as características do GSM. Opera com uma potência de transmissão menor e com as mesmas bandas de *uplink* e *downlink* que o GSM. As células podem ter diferentes tamanhos: picocélulas são mais usadas em ambientes fechados; microcélulas são usadas em áreas rurais e para cobrir estradas e rodovias. Células menores utilizam menos potência para transmissão e permitem maior reutilização de freqüências. A arquitetura poderia ser celular mista no caso de descontinuidade no processo de *handoff* e isso consiste em que essas micro e picocélulas estariam dentro de uma célula convencional. Então, no caso de defeito no sistema, a célula convencional assumiria o processamento.

7.7.3. Comunicações Pessoais Universais (UPT)

UPT é um conceito de serviço que utiliza as capacidades das IN's (*intelligent networks*) e ISDN's (i*ntegrated services digital network*) para promover mobilidade pessoal e portabilidade de serviço para usuários finais. No ambiente de serviço UPT, a cada usuário será atribuído um único número pessoal (número UPT). Tal número será discado por outro

usuário do serviço para alcançar esse usuário UPT, e também será utilizado para identificar um usuário no momento do pedido de um serviço, como, por exemplo, origem da chamada ou modificação do perfil de serviço. Através de modificação do perfil de serviço, o usuário UPT será, não apenas capaz de designar terminais específicos (fixos ou móveis) para receber e originar chamadas, como também será capaz de invocar serviços suplementares por assinatura como chamadas protegidas e prosseguimento de chamadas. Os serviços e características específicos disponíveis para um usuário UPT conforme este se move através de diferentes redes e utiliza diferentes terminais dependerão, é claro, das capacidades dos terminais assim como das redes os servem. As principais características de UPT incluem:

- Mobilidade pessoal baseada em único número pessoal (número UPT);
- Portabilidade e controle de serviço;
- Cobrança baseada em número UPT ao invés de número/identidade do terminal;
- Disponibilidade/acessibilidade de serviço através de múltiplas redes (ISDN, PSTN, PLMN) e diferentes tipos de terminais (fixo, portátil e móvel);
- Conjunto de serviços escolhidos por menu pelo usuário UPT na provisão do serviço;
- Segurança de acesso e localização confidencial através de autenticação e verificação do usuário.

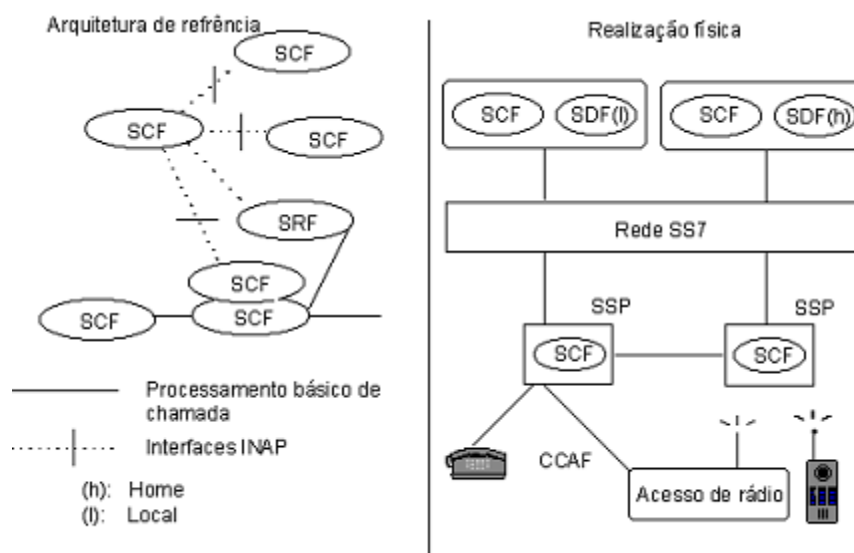



Figura 7-7.1: Arquitetura funcional e possível implementação para UPT

O serviço UPT está sendo especificado em fases. O conjunto de serviços UPT número um (UPT SS1) foi padronizado pelo ITU-T como recomendação F.851, que suporta telefonia

sobre redes PSTN, ISDN e PLMN, junto com um conjunto de características essenciais e opcionais para mobilidade pessoal. O ITU-T desenvolveu também uma arquitetura funcional para UPT baseada na arquitetura para IN. Estas especificações de interface se baseiam no protocolo de aplicação IN (INAP). A arquitetura funcional UPT é apresentada na Figura 7.11 junto com uma possível implementação física.

7.7.4. Sistemas de terceira geração

FPLMTS

O padrão FPLMTS (*Future Public Land Mobile Telecommunications*), que vem sendo desenvolvido pelo ITU-T e pelo ITU-R, tem como intuito formar a base para a terceira geração de sistemas sem fio, que pode acabar com a incompatibilidade que existe atualmente entre os diversos sistemas móveis e consolidá-los em uma infraestrutura de rede e de rádio homogênea, capaz de oferecer um vasto conjunto de serviços de telecomunicações em escala global. O UMTS (*Universal Mobile Telecommunications System*) é um padrão paralelo, que está sendo desenvolvido pelo ETSI e está em conformidade com o FPLMTS.

Algumas das características do sistema para o FPLMTS incluídas no ITU-R são:

- Sistema digital utilizando a banda 1,8–2,2 GHz;
- Múltiplos ambientes de rádio (celular, sem fio, por satélite);
- Terminais multimodo para possibilitar *roaming*;
- Um grande conjunto de serviços de telecomunicações;
- Alta qualidade e integridade, se comparado às redes fixas;
- *Roaming* internacional e possibilidade de *handoff* entre os sistemas;
- Utilização de capacidades de IN para gerenciamento da mobilidade e controle do serviço;
- Níveis altos de segurança e privacidade;
- Arquiteturas de redes abertas e flexíveis.

A Conferência Mundial Administrativa de Rádio de 1992 (WARC'92) reservou uma banda de freqüência comum (1,8–2,2 GHz) para os ambientes de rádio no FPLMTS. No entanto, de acordo com o ITU-R, é necessária uma abordagem diferente dos

desenvolvimentos de transmissão de rádio realizados anteriormente para que se possa maximizar as características comuns através do conjunto de ambientes de rádio no FPLMTS, de forma que terminais de baixo custo e com multimodo possam ser desenvolvidos. O ITU-R está identificando as funções relacionadas que podem ser comuns para múltiplos ambientes operando com rádio, assim como também um conjunto mínimo de funções que depende de cada ambiente individualmente.

Uma arquitetura de rede flexível e aberta foi proposta pelo FPLMTS, o qual permitirá a implementação de sistemas sem fio de terceira geração, tanto como redes isoladas com uma ligação via conexões de *gateways* com as redes públicas, quanto como um sistema integrado onde o gerenciamento de mobilidade e a funcionalidade de controle de chamada são partes dos elementos da rede em uma rede fixa.

No quadro de tempo proposto para implementação do FPLMTS (entre 2000 e 2005), espera-se que recursos e capacidades de IN e SS7 estejam amplamente implantados não apenas nas redes públicas comutadas, mas também em sistemas de segunda geração como o GSM/DCS1800. A principal vantagem da separação entre controle de serviço e controle de chamada/conexão fornecida pela IN está na habilidade potencial de suportar um serviço no topo de qualquer *backbone* que admita os necessários "pontos de disparo" para o controle do serviço. No entanto, em teoria, procedimentos específicos para móveis, como registro de localização, autenticação, *handoff* etc, podem ser suportados por redes IN's estruturadas (fixas ou móveis). O FPLMTS utilizará as capacidades das IN's para fornecer gerenciamento de mobilidade e funções de chamada/conexão, assim como também a facilidade de rápida introdução ao serviço.

O ITU-T está desenvolvendo a arquitetura funcional e os protocolos de sinalização para FPLMTS que utilizarão os recursos das IN's. A arquitetura funcional corrente do ITU-T para o FPLMTS é mostrada na figura abaixo. Ela formará a base para a especificação do acesso de rádio, das interfaces de rede e dos protocolos. Como ilustrado na figura, as funções podem ser particionadas em termos de Funções de Terminais de Rádio, Funções de Acesso de Rádio e Funções de Rede.

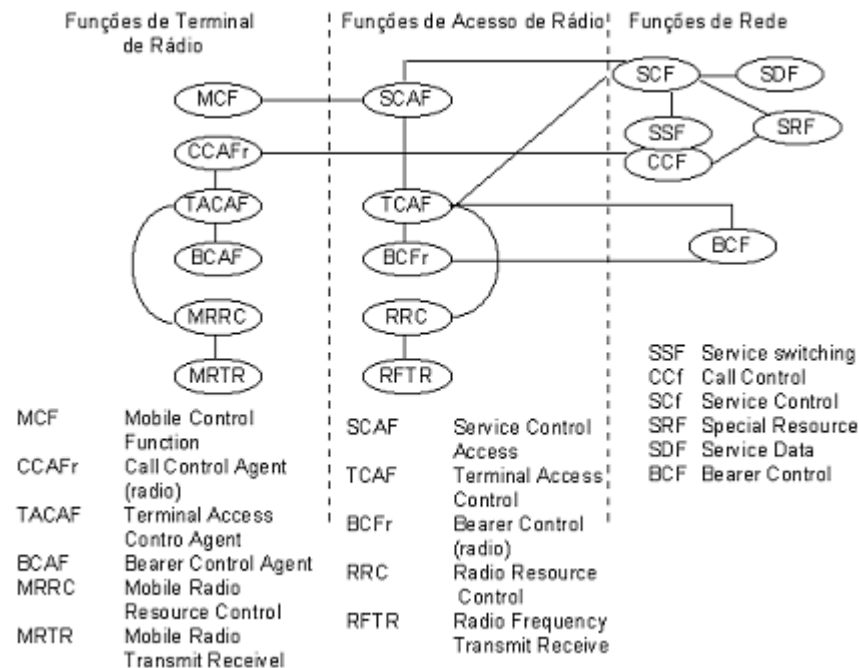



Figura 7-7.2: Arquitetura funcional para o FPLMTS

O FPLMTS e o UMTS foram inicialmente concebidos para prover avanços significativos além da primeira e da segunda gerações de sistemas móveis, representados por celulares analógicos e digitais e por sistemas de telecomunicações sem fio. Essas características avançadas incluem: capacidade para *roaming* ao redor do mundo; um grande conjunto de serviços oferecidos, incluindo transmissão de dados com elevada taxa e uso de serviços audiovisuais; e o uso de um único terminal em diferentes ambientes de operação de rádio. Entretanto, já existe um investimento significativo e crescente em sistemas celulares digitais e de telecomunicações sem fio. De forma a garantir o retorno no investimento em sistemas móveis de segunda geração e que o sistema PCS seja completamente realizado, a transição dos sistemas de segunda para os de terceira geração é suave e dirigida pelo mercado. Estratégias de transição cabíveis e a introdução de cronogramas para FPLMTS/UMTS são necessárias.

UMTS

Na Europa, o objetivo em longo prazo é um sistema de comunicações móvel e universal (UMTS – *Universal Mobile Telecommunications Systems*), que unifique todos os sistemas celulares, sem fio, RLL, LAN's sem fio, rádio móvel privado (PMR) e *paging*. A idéia é fornecer o mesmo tipo de serviço por toda a parte. A única limitação seria a taxa de

transmissão disponível, que dependeria da localização e da carga do sistema. O escopo é um sistema com múltiplos operadores com várias arquiteturas de células misturadas e suporte para multimídia. Esses requisitos são difíceis de serem atingidos, pois possuem implicação direta com a interface de rádio e a estrutura de protocolos.

O GSM e o DCS 1800 estão atualmente sendo levados na direção do UMTS. Os serviços de malha local de rádio para os lares e escritórios, assim como funcionalidades de PBX nos comutadores públicos, poderiam assumir funções sem fio se o serviço fosse fornecido a um preço adequado. Um tópico ainda em discussão diz respeito às altas taxas de transmissão de dados. Sob condições otimistas, um esquema de modulação alternativo e a utilização de alocações assimétricas de *slots* podem subir a taxa para no máximo 64 kbps, mas certamente não para 2 Mbps. A eficiência do espectro de freqüência e o baixo custo para usuários finais são aspectos importantes ainda em discussão.

O DECT (Digital Enhanced Cordless Telecommunications) poderia crescer além das áreas atuais de aplicações, como telefonia sem fio, PBX e Teleponto. Estudos estão sendo realizados atualmente para troca da malha local. A extensão dos limites através de sistemas de antenas mais avançados e alguma forma de equalização, assim como repetidores, são questões importantes neste contexto. O DECT poderia suportar o mercado de massa a baixo custo em áreas densamente povoadas, deixando a cobertura em áreas extensas para sistemas celulares (UMTS parcial). A alocação dinâmica de canal e o alto grau de flexibilidade no fornecimento de novos serviços, pela combinação de canais para obter taxas de transmissão de dados elevadas (sem modificação do padrão, talvez com transceptores de até 200 kbps) são outros trunfos. Finalmente, o uso da interface aérea do DECT em conjunto com a infraestrutura do GSM está em estudo.

Os projetos ATDMA (*Advanced TDMA*) e o CODIT (*Code Division Testbed*), do programa R&D em tecnologias de comunicação avançadas na Europa (RACE), seguem abordagens revolucionárias para o UMTS. O primeiro projeto utiliza TDMA como principal técnica de acesso e o segundo projeto utiliza o CDMA. Ambos buscam altas taxas de transmissão. Nos testes, taxas acima de 64 kbps e de 128kbps estão sendo conseguidas, incluindo ainda taxas de rajadas acima de 2 Mbps. Também estão em estudo o acesso à Internet e a transmissão de pacotes sem conexão com pequeno atraso.

Assim, a principal linha de pesquisa na Europa está explorando o potencial do GSM e do DECT, incluindo evoluções posteriores e a interligação entre os padrões. Outros padrões estão sendo finalizados para LAN's sem fio (HIPERLAN) e PMR (TETRA, de Sistema de

Rádio Trans Europeu com Troncos), por exemplo. Concomitantemente, a definição e a especificação do UMTS está sendo realizada dentro do ETSI e suportado pelo RACE.

7.7.5. WCDMA

O WCDMA (*Wideband CDMA*), padrão norte-americano conhecido por IS-665, foi legalizado em junho de 1995. Possui uma largura de banda com 5, 10 ou 15 MHz para a comunicação pessoal já estabelecida. Por adotar 32 kbps ADPCM (*Adaptive Differential Pulse Code Modulation*), técnica de codificação de voz muito eficiente, o WCDMA é um sistema de comunicação móvel pessoal que permite comunicação em movimento de até 100 km/h com alta qualidade.

O WCDMA é um sistema de transmissão de dados sem fio de banda larga (*wideband)*, baseado no método de propagação de espectro, e possui um importante papel na evolução das comunicações móveis sem fio, porque tem quase a mesma performance de transmissão que as comunicações com fio. Já é adotado como o sistema de comunicação da terceira geração em várias comunidades de padronização, na Europa e no Japão. A velocidade do tráfego de dados chega a 2 Mbps para cada usuário.

Serviços com comutação por pacotes e por circuitos podem ser livremente combinados, com larguras de banda variáveis, e serem entregues simultaneamente ao usuário com níveis de qualidade específicos. As exigências da largura de banda para um usuário podem ser mudadas durante uma sessão.

O WCDMA pode ser desenvolvido dentro da nova faixa de freqüência de 2 GHz, reservada para serviços da terceira geração UMTS/IMT-2000, ou dentro das faixas de freqüência já utilizadas. Esse padrão pode ser considerado como uma tecnologia com baixos riscos, o que é uma consideração importante, tamanha a velocidade com que os mercados estão se movimentando.

Usando WCDMA como o padrão para UMTS na Europa e para IMT-2000 globalmente, procura-se realizar as esperanças e as intenções originais da padronização da próxima geração: pela primeira vez, ajustar um padrão verdadeiramente global para comunicações sem fio, conduzindo à oportunidade de *roaming* global.

7.7.6. Quarta geração

A quarta geração ainda é apenas um conceito, pois não existe nenhum padrão em desenvolvimento. Nem sequer há uma definição mundialmente aceita do que ela deverá ser. Isso não quer dizer, entretanto, que não tenha havido avanços significativos rumo à 4G nos últimos anos.

Já existem pontos de consenso relevantes. Um deles se refere, por exemplo, à possibilidade da 4G promover unificação das diferentes redes sem fio, incluindo as tecnologias de redes locais como a Wi-Fi (*Wireless Fidelity*), que outros preferem chamar simplesmente de *Wireless* LAN, ou ainda pela denominação dada pelo IEEE, que a designa por IEEE 802.11 (com as versões a, b, c, d, e, f e g). Além dessas, é preciso mencionar as redes Hiper LAN/2 e HomeRF.

A 4G deverá, por fim, integrar as chamadas redes pessoais, como a rede Bluetooth, para curtas distâncias. Sob o guarda-chuva da 4G, haverá, então, a interligação de todos os tipos possíveis de aparelhos, dispositivos ou equipamento móveis, interligando os diversos tipos de redes – algumas delas utilizando espectro licenciado e outras, não.

Em resumo, a 4G tende a significar que "tudo poderá funcionar em qualquer lugar". Os especialistas que trabalham no desenvolvimento da 4G imaginam que a nova rede deva ser capaz de interagir em especial com o conteúdo de informação baseada na internet.

A internet tem acelerado o interesse por essa computação sem limites. As redes sem fio podem viabilizar esse interesse. Os organismos incumbidos da padronização da internet estão também explorando conceitos de apoio, incluindo a linguagem XML, como sucessora da HTML. Vale lembrar também que a condição básica de interoperabilidade da internet tem sido a adoção de padrões universais, tais como o protocolo IP, a linguagem HTML, procedimentos e interfaces. Do mesmo modo, a integração das redes móveis de 4G com a internet dependerá da adoção de características universais.

Apêndice A

TEORIA DAS FILAS

A.1. INTRODUÇÃO

Ao efetuarmos certos tipos de estudos de planejamento, é comum depararmos com problemas de dimensionamento ou fluxo cuja solução é aparentemente complexa. O cenário pode ser uma fábrica, o trânsito de uma cidade, um escritório, um porto, uma mineração, etc. Geralmente estamos interessados em saber:

- Qual a quantidade correta de equipamentos (sejam eles máquinas, veículos, etc.);
- Qual o melhor *layout* e o melhor fluxo dentro do sistema que está sendo analisado.

Ou seja, desejamos que nosso sistema tenha um funcionamento eficiente ou otimizado. Por **otimizado** queremos dizer que teremos um **custo adequado** e que teremos **usuários satisfeitos** com o ambiente ou com o serviço oferecido. Chamamos tais estudos de **modelagem de sistemas**.

Estudos de modelagem de sistemas podem envolver modificações de *layout*, ampliações de fábricas, troca de equipamentos, reengenharia, automatização, dimensionamento de uma nova fábrica, etc. Assim, dado um determinado objetivo de produção ou de qualidade de atendimento, o estudo vai procurar definir a quantidade de atendentes (equipamentos, veículos, pessoas, etc.) que devem ser colocados em cada estação de trabalho, assim como o melhor *layout* e o melhor fluxo. Para dimensionar adequadamente um sistema devemos dedicar especial atenção aos chamados **gargalos**, ou seja, os pontos onde ocorrem filas. Dizemos também que um sistema ou processo adequadamente dimensionado está **balanceado**.

Qualquer pessoa sabe exatamente o que são filas em decorrência das experiências que o dia-a-dia nos coloca. Nós entramos em uma fila para descontar um cheque em um banco, para pagar pelas compras em um supermercado, para comprar ingresso em um cinema, para pagar o pedágio em uma estrada e em tantas outras situações. Em engenharia, também existem situações de fila como programas esperando para serem atendidos pelo

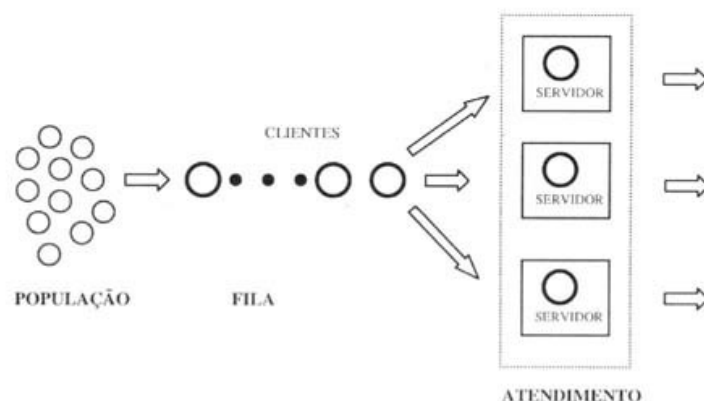
microprocessador, ou para ter permissão de acesso a um dispositivo de entrada e saída, assim como assinantes esperando por uma linha disponível para efetuar uma ligação telefônica.

Além de não serem simpáticas, as filas têm ainda o lado desfavorável do custo. Isto é válido em qualquer ambiente, indo de fábricas a um supermercado. Por exemplo, nas fábricas a existência de fila em um equipamento pode implicar espera por peças que necessitam ser processadas, o que ocasiona um aumento nos tempos de produção. As conseqüências disto são aumento nos custos e atrasos no atendimento aos pedidos dos clientes.

O ideal seria a não existência de filas. Se isto realmente ocorresse, certamente não teríamos clientes aborrecidos, mas isto nem sempre é possível, pois tal situação implicaria em um custo extremamente elevado para manter uma disponibilidade infinita de recursos para atender a qualquer demanda de clientes. Portanto, apesar de não serem simpáticas e causarem prejuízos, temos que conviver com as filas na vida real, visto ser antieconômico superdimensionar um sistema para que nunca existam filas. O que se tenta obter é um balanceamento adequando que permita um atendimento aceitável pelo menor custo.

A abordagem matemática de filas se iniciou no princípio do século XX (1908) em Copenhague, Dinamarca, através de A. K. Erlang, considerado o pai da Teoria das Filas, quando trabalhava em uma companhia telefônica estudando o problema de redimensionamento de centrais telefônicas. Foi somente a partir da segunda guerra mundial que a teoria foi aplicada a outros problemas de filas. Apesar do enorme progresso alcançado pela teoria, inúmeros problemas não estão ainda adequadamente resolvidos por causa da complexidade matemática.

A.2. CONCEITOS BÁSICOS

Na figura abaixo aparecem os elementos que compõem uma fila. Nela temos que, de uma certa **população**, surgem **clientes** que formam uma **fila** e que aguardam por algum tipo de **serviço**. O termo cliente é usado de uma forma genérica e pode designar tanto uma pessoa como uma máquina. Como um sinônimo de cliente usa-se também o termo **transação** ou **entidade**. O **atendimento** é constituído de um ou mais **servidores** (que podem ser também chamados de **atendentes** ou de **canais de serviços**) que realizam o atendimento dos clientes.

Antes de observar o funcionamento de uma fila, é melhor primeiro conceituar alguns termos da Teoria das Filas:

- **Clientes e Tamanho da População:** Um cliente é proveniente de uma população. Quando a população é muito grande, a chegada de um novo cliente a uma fila não afeta a taxa de chegada de clientes subseqüentes. Portanto, as chegadas se tornam independentes.

- **Processo de Chegada:** Considere como exemplo um posto de pedágio com cinco atendentes. Podemos constatar, por exemplo, que o processo de chegada entre 7 e 8 horas da manhã pode ser definido por 20 automóveis por minuto ou 1 automóvel a cada 3 segundos. Trata-se de um valor médio, pois não significa que em todo intervalo de 1 minuto chegarão 20 automóveis. Em alguns intervalos de 1 minuto pode-se constatar a chegada de 10, 15, 25 ou até mesmo 30 automóveis. Conseqüentemente o intervalo entre as chegadas não é rígido. O número fornecido de 3 segundos representa, assim, o intervalo médio entre chegadas no período de 7 às 8 horas da manhã. Quando se estudam filas, o ritmo de chegada é uma importante variável aleatória. Para quantificar esta variável usa-se a letra grega λ , que simboliza o ritmo de chegada e se usa IC para o intervalo médio entre chegadas. Assim, no exemplo dado:

$$\lambda = 20 \text{ carros por minuto.}$$
$$IC = 3 \text{ segundos.}$$

- **Processo de Atendimento:** Continuando no exemplo do pedágio e observando um atendente em serviço, podemos constatar, por exemplo, que ele atende 6 veículos por

minuto ou que gasta 10 segundos para atender um veículo. Esses valores são médios e, para descrevê-los corretamente, devemos também lançar mão da distribuição de probabilidades. O processo de atendimento também é uma importante variável aleatória. A letra grega μ é usada para significar ritmo de atendimento e TA para designar tempo ou duração do serviço ou atendimento. Assim, no exemplo dado teremos:

μ = 6 clientes por minuto.

TA = 10 segundos por cliente.

- **Número de Servidores:** O mais simples sistema de filas é aquele de um único servidor que pode atender um único cliente de cada vez. Conforme aumente o ritmo de chegada, podemos manter a qualidade do serviço aumentando convenientemente o número de servidores. Esta é, portanto, uma das características de uma fila que podemos utilizar para modelar um sistema de filas.

- **Disciplina da Fila:** Trata-se da regra que define qual o próximo a ser atendido e o comum é que o primeiro da fila é atendido ou, de uma maneira mais ampla, "o primeiro a chegar é o primeiro a ser atendido" (em inglês, diz-se FIFO: *First in First Out*). Outras disciplinas podem existir tais como "último a chegar é o primeiro a ser atendido" (em inglês, LIFO: *Last in First Out*), serviço por ordem de prioridade, serviço randômico, etc.

- **Tamanho Médio da Fila:** Esta é a característica que mais consideramos ao nos defrontarmos com a opção de escolher uma fila. O ideal é a situação de chegar e ser atendido imediatamente (fila zero). Quando a fila é de um tamanho razoável, intuitivamente sabemos que o tempo de espera na fila será longo. O tamanho da fila não é constante e quando os ritmos de chegada e de atendimento são constantes, o tamanho da fila oscila em torno de um valor médio.

- **Tamanho Máximo da Fila:** Quando os clientes devem esperar, alguma área de espera deve existir (por exemplo, uma determinada quantidade de um *buffer* de memória). Observa-se, na vida real, que os sistemas existentes são dimensionados para uma certa quantidade máxima de clientes em espera, sendo este dimensionamento geralmente feito com base em uma experiência real. Quando existe um crescimento na demanda, se faz

uma ampliação também baseada na experiência com o manuseio do referido sistema. Observam-se, também, casos em que um novo cliente que chega pode ser recusado, devendo tentar novamente em um outro instante (por exemplo, a tentativa de conseguir uma linha telefônica recebendo o sinal de "ocupado" ou de que não há linha disponível).

- **Tempo Médio de Espera na Fila:** Esta é a característica capaz de nos causar irritação quando estamos em uma fila de espera. O ideal é que não exista espera, mas esta nem sempre é a melhor solução do ponto de vista econômico. Se entrarmos em uma fila com 10 pessoas à nossa frente, o tempo de espera será igual ao somatório dos tempos de atendimento de cada um dos clientes na nossa frente ou, possivelmente, será igual a 10 vezes a duração média de atendimento. Tal como o tamanho médio da fila, o tempo médio de espera depende dos processos de chegada e de atendimento.

Como exemplo de aplicação dessas características de uma fila, imagine-se agora instalado em uma poltrona dentro de um banco, com a finalidade de observar o funcionamento da fila formada por pessoas que desejam um novo talão de cheques. No período de meia hora você verificou que chegaram ao sistema 12 pessoas:

Cliente	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Intervalo	2	3	3	3	5	0	1	5	1	4	1	2
Momento	2	5	8	11	16	16	17	22	23	27	28	30

O valor zero acima significa que o sexto cliente chegou junto com o quinto. O valor médio dos dados cima é de 2,5 minutos e, portanto, o sistema acima funcionou com um ritmo médio de 24 chegadas por hora.

Por outro lado, os dados anotados para cada atendimento são os seguintes (com os valores dados em minutos):

Cliente	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Duração	1	2	1	1	3	2	1	4	2	3	1	3

O valor médio dos dados acima é de 2,0 minutos e, portanto, podemos dizer que o servidor tem uma capacidade de atender 30 clientes por hora. Assim o tempo em fila de cada um dos clientes é dado por:

Cliente	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Tempo em Fila	0	0	0	0	0	3	4	0	3	1	3	2

Portanto, analisando o gráfico da figura a seguir, podemos concluir:

- Tempo Médio na Fila = (3+4+3+1+3+2)/12 = 16/12 = 1,33 minuto.
- Tamanho Médio na Fila = (3+4+3+1+3+2)/35 = 16/35 = 0,46 cliente.

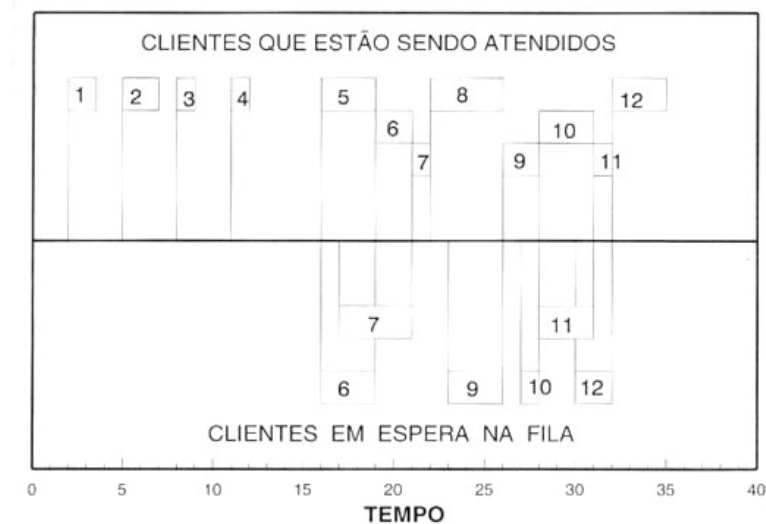



Figura A-2.2: Funcionamento da Fila do posto Bancário

Imagine agora que o mesmo problema fosse proposto ao leitor da seguinte forma: clientes chegam a um posto de serviço a um ritmo de 24 chegadas por hora (ou um cliente a cada 2,5 minutos) e são atendidos por um servidor capaz de atender 30 clientes por hora (ou 2,0 minutos para cada cliente). A pergunta que se faz é: haverá fila?

De posse desses dados, a nossa tendência é inferir que o sistema se comportará de tal forma que tanto o processo de chegada como o de atendimento são regulares e, portanto, não haverá formação de filas. Processos como este são raros na vida real. Como se pode deduzir, a existência de filas ocorre em decorrência do fato de que os processos não são regulares e a aleatoriedade tanto ocasiona filas como ocasiona também longos períodos de inatividade para o servidor. Isso ocorre porque em um determinado instante podem chegar mais clientes que a

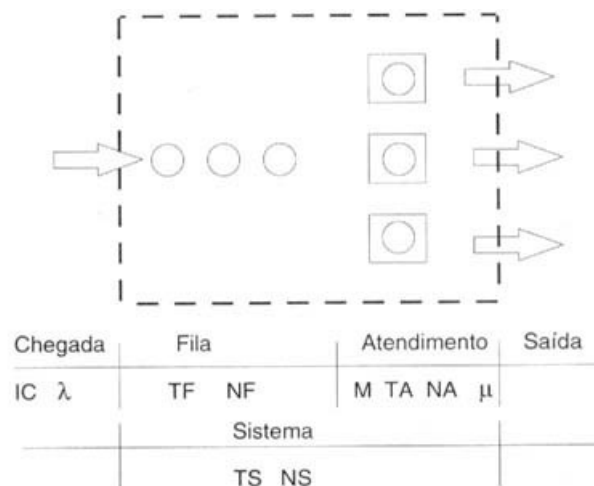
capacidade de atendimento daquele momento, gerando filas temporárias ou o atendimento de um determinado cliente pode ser maior que a média, fazendo com que os clientes que chegam depois fiquem em uma fila.

A abordagem matemática de filas pela Teoria das Filas exige que exista estabilidade no fluxo de chegada e no processo de atendimento, ou seja, os valores de λ e μ se mantém constantes no tempo. Por exemplo, observando o tráfego telefônico, verifica-se que em determinados horários do dia há mais demanda por ligações do que em outros horários, ou seja, o fluxo de chegada varia conforme o horário do dia. Não existe estabilidade para o ritmo de chegada no período de um dia, e, portanto, não podemos analisar seu funcionamento pela Teoria das Filas, a menos que usemos alguns artifícios como, por exemplo, retalhar o período global em períodos parciais. Infelizmente, isto torna mais complexa a abordagem pela Teoria das Filas.

Outra exigência para que o processo seja estável é que os servidores sejam capazes de atender ao fluxo de chegada. No caso de uma fila e um servidor, isso significa dizer que obrigatoriamente $\mu > \lambda$ (a capacidade de atendimento é maior que o ritmo de chegada) e, caso isto não ocorra, resulta que o tamanho da fila aumentará infinitamente.

Em sistemas estáveis, todas as características aleatórias das filas se mantêm estáveis o tempo todo, significando que oscilam em torno de um valor médio como, por exemplo, o tamanho médio, o tempo médio de espera, o tempo médio de atendimento, etc.

A.3. PRELIMINARES MATEMÁTICAS

Consideremos o sistema de filas da Figura A.3.1, em situação estável, na qual clientes chegam e entram em fila, existindo M servidores para atendê-los. Seja λ o ritmo médio de chegada e μ o ritmo médio de atendimento de cada atendente. Dentre as variáveis aleatórias que estudaremos neste Apêndice, algumas serão freqüentemente citadas e as chamaremos *variáveis aleatórias fundamentais*. É o que mostramos na Figura A.3.1 e explicaremos a seguir.

- Variáveis Referentes ao Sistema:
  - TS →Tempo Médio de Permanência no Sistema.
  - NS →Número Médio de Clientes no Sistema.

- Variáveis Referentes ao Processo de Chegada:
  - λ →Ritmo Médio de Chegada.
  - IC →Intervalo Médio entre Chegadas.

  Por definição: $IC = 1/\lambda$.

- Variáveis Referentes à Fila:
  - TF →Tempo Médio de Permanência na Fila.
  - NF →Número Médio de Clientes na Fila.

- Variáveis Referentes ao Processo de Atendimento:
  - TA →Tempo Médio de Atendimento ou de Serviço.
  - M →Quantidade de Atendentes.
  - NA →Número Médio de Clientes que estão sendo Atendidos.
  - μ →Ritmo Médio de Atendimento de cada Atendente.

  Por definição: $TA = 1/\mu$.

Existem duas relações óbvias entre as variáveis aleatórias mostradas na Figura A.3.1:

$$NS = NF + NA. \quad (A.1)$$

$$TS = TF + TA. \quad (A.2)$$

Pode-se também chegar à conclusão de que:

$$NA = \frac{\lambda}{\mu} = \frac{TA}{IC}. \quad (A.3)$$

Portanto:

$$NS = NF + NA = NF + \frac{\lambda}{\mu} = NF + \frac{TA}{IC}. \quad (A.4)$$

Para o caso particular de uma fila e um atendente, chamamos de taxa de utilização do atendente à expressão:

$$\rho = \frac{\lambda}{\mu}. \quad (A.5)$$

Que é exatamente o número médio de clientes que estão sendo atendidos pelo servidor.

Para o caso mais geral de um sistema com M servidores, a expressão se torna:

$$\rho = \frac{\lambda}{M\mu}. \quad (A.6)$$

Assim, ρ representa a fração média do tempo em que cada servidor está ocupado. Por exemplo, com um atendente, se chegam 4 clientes por hora e se o atendente tem capacidade para atender 10 clientes por hora, dizemos que a taxa de utilização é de 0,4 e podemos também afirmar que o atendente fica 40% do tempo ocupado e 60% do tempo livre (essa afirmativa é intuitiva, mas pode ser matematicamente demonstrada).

Como serão estudados apenas os sistemas estáveis (os atendentes sempre serão capazes de atender ao fluxo de chegada) teremos sempre que $\rho < 1$. Quando $\rho = 1$, o atendente trabalhará 100% do tempo.

Um outro conceito importante na Teoria das Filas é o de tráfego. Define-se a intensidade de tráfego através da expressão:

$$i = \left\lceil \frac{\lambda}{\mu} \right\rceil = \left\lceil \frac{TA}{IC} \right\rceil. \quad (A.7)$$

Onde o símbolo $\lceil\ \rceil$ designa o menor inteiro maior que o número real considerado.

A intensidade de tráfego é medida em *erlangs* em homenagem a A. K. Erlang. Na prática, *i* representa o número mínimo de servidores necessário para atender dado fluxo de tráfego. Por exemplo, se $\lambda = 10$ clientes/hora e TA = 3 minutos (ou $\mu = 20$ clientes/hora) temos que $\lambda/\mu = 0{,}5$ ou $i = 1$, e concluímos dizendo que 1 atendente é suficiente para o caso. Se o fluxo de chegada aumentar para $\lambda = 50$ clientes/hora, temos que $\lambda/\mu = 2{,}5$ ou $i = 3$, isto é, há necessidade de, no mínimo, 3 atendentes. Na indústria telefônica esta variável é bastante utilizada ao se referir a tráfego em troncos telefônicos.

J. D. C. Little demonstrou que, para um sistema estável de filas, temos:

$$NF = \lambda\ TF. \quad (A.8)$$

$$NS = \lambda\ TS. \quad (A.9)$$

Estas fórmulas são muito importantes, pois, assim como as equações (A.1) e (A.2), fazem referência a quatro das mais importantes variáveis aleatórias de um sistema de filas: NS, NF, TS e TF. Por exemplo, se além de λ e μ conhecemos TS, podemos obter as outras variáveis aleatórias assim:

$$NS = \lambda\ TS; \quad (A.10)$$

$$TA = \frac{1}{\mu}; \quad (A.11)$$

$$TF = TS - TA; \quad (A.12)$$

$$NF = \lambda\ TF. \quad (A.13)$$

É importante salientar que todos os resultados acima independem da quantidade de servidores e do modelo de fila, pois se tratam de relações fundamentais básicas.

Apresentamos a seguir os postulados básicos que se aplicam a quaisquer sistemas de filas nos quais existe estabilidade, ou seja, λ é menor que μ em todas as estações de trabalho (o ritmo médio de chegada é menor que o ritmo médio de atendimento).

- Em qualquer sistema estável, o fluxo que entra é igual ao fluxo que sai:

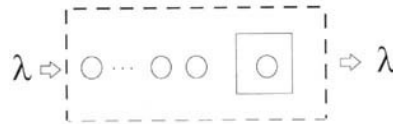



Figura A-3.2: Sistema em que o fluxo de entrada é igual ao fluxo de saída

- Em um sistema estável, o fluxo de entrada se mantém nas diversas seções do sistema:

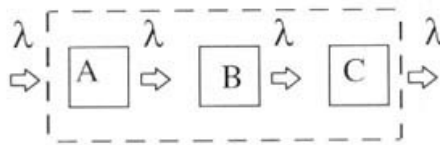



Figura A-3.3: Fluxo entre subdivisões de um sistema

- Em um sistema estável, a junção de fluxos equivale às suas somas:

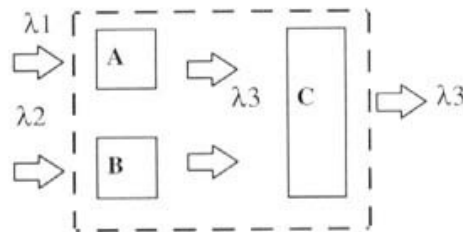
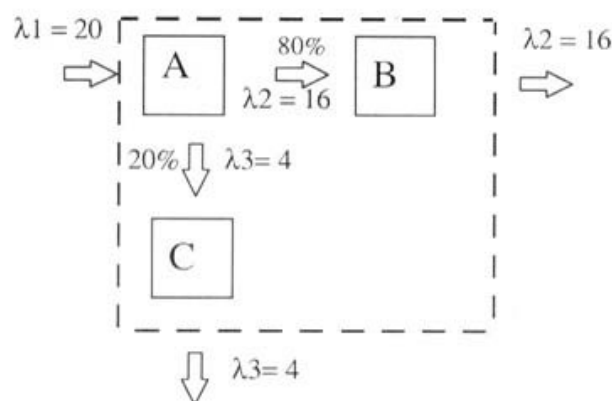



Figura A-3.4: Junção de dois fluxos de entrada, onde $\lambda_1 + \lambda_2 = \lambda_3$

- Em um sistema estável, a soma dos fluxos que saem é igual à soma dos fluxos que entram em um determinado sistema:

A.4. O PROCESSO DE CHEGADA

Nesta seção será apresentado o processo de chegada quantitativamente sob a ótica da teoria das probabilidades e da estatística.

Para exemplificar, vamos nos basear nos dados da tabela a seguir, que mostra anotações sobre a quantidade de chamadas telefônicas feitas em uma pequena central. Os valores da tabela mostram quantos assinantes fizeram uma ligação a cada intervalo de 1 minuto entre 7 e 8 horas da manhã. Vemos, por exemplo, que, no primeiro, minuto ligaram 2 assinantes e que, no segundo minuto, ligou 1 assinante.

Ritmos de Chamadas Telefônicas									
2	1	2	1	0	2	1	0	1	2
0	2	3	1	3	1	3	4	5	1
2	0	1	2	1	0	1	1	0	2
2	2	3	2	2	3	2	3	3	2
1	6	0	2	3	7	0	2	2	0
4	1	1	1	1	8	4	3	1	4

Tabela A-4.1: Chamadas telefônicas realizadas em cada minuto entre 7 e 8 horas

Nas 60 anotações da tabela acima foram feitas 120 chamadas telefônicas, o que nos fornece λ = 2 chamadas/minuto.

Quando trabalhamos com dados tais como os acima, devemos nos valer da Estatística para analisá-los corretamente, pois desejamos conhecer não apenas o valor médio, o valor mínimo e o valor máximo como também desejamos saber como os valores se distribuem em torno da média. Para efetuar uma análise estatística destes dados devemos começar agrupando-os como, por exemplo, na tabela abaixo. Observe que a coluna *freqüência relativa*

permite uma análise mais adequada dos dados do que a coluna *freqüência absoluta*. Por exemplo, vemos que a freqüência absoluta da ocorrência de 3 chamadas telefônicas foi de 9. No entanto, essa informação ainda necessita da consideração de que foram 9 ocorrências de três chamadas em um minuto em 60 amostras. No entanto, usando a freqüência relativa, podemos afirmar que: em 15% das amostras houve a ocorrência de três chamadas telefônicas em um minuto. O que é uma informação mais completa do ponto de vista estatístico. Caso o experimento pudesse ser realizado um número infinito de vezes, a freqüência relativa iria tender para a probabilidade da ocorrência de três chamadas telefônicas em um minuto.

Ritmo	Freqüência Absoluta	Freqüência Relativa
0	9	0,150
1	17	0,283
2	17	0,283
3	9	0,150
4	4	0,066
5	1	0,017
6	1	0,017
7	1	0,017
8	1	0,017
9	0	0
10	0	0

Tabela A-4.2: Ritmo x Freqüência

Agora fazemos a seguinte pergunta: qual é a distribuição estatística que melhor se aproxima dos dados reais acima? Para respondermos a esta pergunta necessitamos comparar os dados das freqüências relativas obtidas na Tabela A-4.2 com as distribuições conhecidas. A Estatística possui uma série de métodos para determinar qual modelo estatístico consegue descrever melhor um conjunto de dados experimentais. Tais métodos são baseados na minimização do erro médio quadrático entre os resultados da distribuição teórica e a distribuição dos dados obtidos por meio de experimentos. Segundo esses métodos, determinou-se que em problemas de chegada em uma fila, a distribuição que melhor descreve o comportamento estatístico é a distribuição de Poisson:

$$f(\mathrm{x}) = \frac{\alpha^{\mathrm{x}} \cdot e^{-\alpha}}{\mathrm{x}!}. \qquad \text{(A.14)}$$

Onde $f(\mathrm{x})$ representa a probabilidade de ocorrerem x chamadas telefônicas.

Calculando a média da distribuição de Poisson, temos que:

$$E[x] = \sum_{x=0}^{\infty} x \cdot f(x) = \sum_{x=0}^{\infty} x \cdot \frac{\alpha^{x} \cdot e^{-\alpha}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\alpha^{x} \cdot e^{-\alpha}}{x!} = \alpha\, e^{-\alpha} \sum_{x=0}^{\infty} \frac{\alpha^{x-1}}{(x-1)!}. \quad (A.15)$$

Fazendo uma substituição de variáveis: k = x+1, temos:

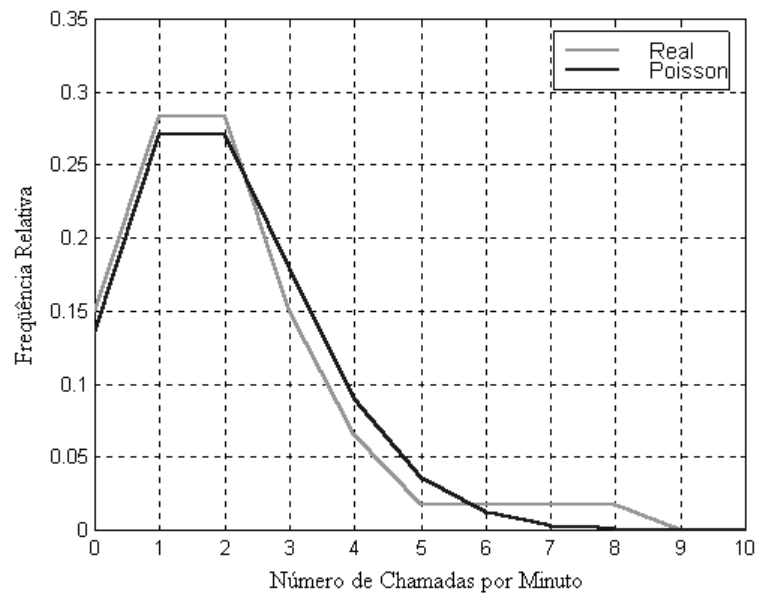
$$E[x] = \alpha\, e^{-\alpha} \sum_{x=0}^{\infty} \frac{\alpha^{x-1}}{(x-1)!} = \alpha\, e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^{k}}{k!}. \quad (A.16)$$

Reconhecendo que o somatório de (A.16) corresponde à série de Taylor da função exponencial, teremos que:

$$E[x] = \alpha\, e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^{k}}{k!} = \alpha\, e^{-\alpha}\, e^{\alpha} = \alpha. \quad (A.17)$$

Considerando que o número médio de chamadas telefônicas em um dado intervalo de tempo *t* pode ser obtido por $\lambda\ t$ – pois λ indica o número médio de chamadas por unidade de tempo – então teremos que $\alpha = \lambda\ t$. Dessa forma, a distribuição de probabilidades do número de chamadas telefônicas em um dado intervalo de tempo dado o ritmo médio de chegada λ, é dada por:

$$f(x) = \frac{(\lambda t)^{x} \cdot e^{-\lambda t}}{x!}. \quad (A.18)$$

Colocando em um mesmo gráfico os valores das freqüências relativas obtidas na Tabela A-4.2 e os valores das probabilidades obtidas através da distribuição (A.18) com λ = 2 chamadas/minuto e t = 1 minuto, obtemos o gráfico da figura a seguir. Nesse gráfico verifica-se que as chamadas telefônicas realmente apresentam uma distribuição muito semelhante à distribuição de Poisson.

Já na figura abaixo são apresentadas diferentes curvas da distribuição de Poisson para diferentes valores de λ , mantendo-se t = 1 minuto. Nota-se claramente que ao deslocar a média, o pico da curva também é deslocado.

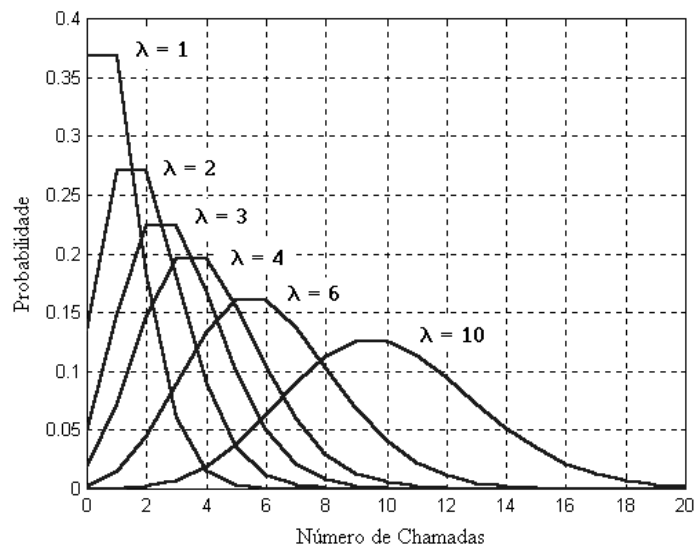



Figura A-4.2: Gráficos de diferentes distribuições de Poisson

Tendo apresentado a distribuição de probabilidades do número de chamadas telefônicas em um intervalo de tempo, consideraremos agora a distribuição de probabilidades do intervalo de tempo entre duas chamadas consecutivas. È importante salientar que o número de chamadas telefônicas em um intervalo de tempo é uma variável aleatória discreta, por isso a distribuição empregada (Poisson) é discreta e o valor de (A.18) para cada x é a

probabilidade da ocorrência de x. No entanto, o tempo entre chamadas é uma variável aleatória contínua, por isso usaremos a função de distribuição e a função de densidade de probabilidades.

A partir de (A.18) podemos calcular a probabilidade de não haver nenhuma chamada telefônica um intervalo de tempo *t*:

$$P[x = 0] = \frac{(\lambda t)^0 \cdot e^{-\lambda t}}{0!} = e^{-\lambda t}. \tag{A.19}$$

De acordo com (A.19), a probabilidade de não haver nenhuma chamada em um intervalo de tempo *t* cai exponencialmente com o aumento do intervalo. Tal resultado é bastante intuitivo, já que é de se esperar que quanto maior o intervalo de tempo menor é a chance de que não ocorra nenhuma chamada telefônica.

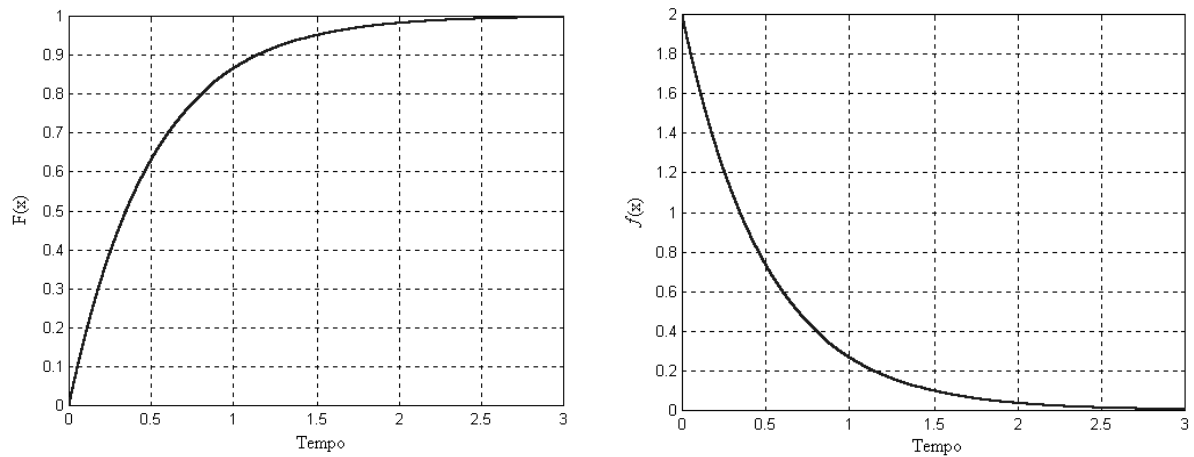
A função de distribuição de probabilidades é obtida através da probabilidade do tempo entre duas chamadas ser menor ou igual a *t*. De acordo com (A.19), obtém-se que:

$$P[IC < t] = 1 - e^{-\lambda t}. \tag{A.20}$$

Então a função de distribuição e a função de densidade de probabilidades do tempo entre duas chamadas sucessivas são dadas, respectivamente, por:

$$F_{IC}(t) = 1 - e^{-\lambda t} \tag{A.21}$$

$$f_{IC}(t) = \lambda\, e^{-\lambda t}. \tag{A.22}$$

A partir de (A.21) e (A.22), nota-se que o intervalo entre duas chamadas consecutivas que seguem a distribuição de Poisson apresentam uma distribuição exponencial de probabilidades; um resultado bem conhecido a partir da teoria de probabilidades.

Na Figura A-4.3 é apresentado o gráfico da função de distribuição exponencial com λ =2 e, na Figura A-4.4, é apresentado o gráfico da respectiva função densidade de probabilidade.

A.5. O PROCESSO DE ATENDIMENTO

De acordo com os estudos de Erlang de 1908, a duração de uma ligação telefônica por cliente é uma variável aleatória que segue aproximadamente uma distribuição exponencial tal qual apresentado em (A.21) e (A.22). No entanto, para o caso do tempo de atendimento, as funções de distribuição e de densidade de probabilidade são dadas por:

$$F_{TA}(t) = 1 - e^{-\mu t} \tag{A.23}$$

$$f_{TA}(t) = \mu\, e^{-\mu t}\,. \tag{A.24}$$

No entanto, é importante considerar que na Teoria das Filas aplicada a outras situações práticas diferentes de telefonia nem sempre o tempo de atendimento segue uma distribuição próxima da exponencial como considerado aqui.

A.6. MODELOS DE FILAS

Na prática, vemos que, em cada caso, temos diferentes tipos de filas. Assim, não podemos considerar o mesmo modelo para todas as filas. No caso prático, um sistema teórico em que o ritmo de chegada segue a distribuição de Poisson e o ritmo de saída segue a distribuição exponencial negativa seria útil somente em casos raros e isolados. Em situações reais, utilizamos diversos modelos de filas que se aproximam mais do caso estudado.

Para descrevermos os modelos de filas, utilizamos a notação **A/B/c/K/m/Z**, onde:

- **A** descreve a distribuição dos intervalos entre chegadas;
- **B** descreve a distribuição do tempo de serviço;
- **c** é a quantidade de atendentes;
- **K** é a capacidade máxima do sistema (número máximo de clientes no sistema);
- **m** é o tamanho da população que fornece clientes;
- **Z** é a disciplina da fila.

Essa notação é chamada de Notação Kendall, em homenagem a seu criador, David Kendall. Os valores de **A** e **B** dependem do tipo de distribuição a que se referem:

- **M:** Exponencial Negativa (Markoviana/Poisson)
- **Em:** Erlang de grau m
- **Hm:** Hiper-exponencial de grau m
- **Determinística**
- **Geral**

Assim, por exemplo, **M/E2/5/20/∞/Randômico** refere-se a uma fila que tem o ritmo de chegada com distribuição Markoviana (ou Poisson), ritmo de atendimento com distribuição Erlang de segundo grau, cinco atendentes, capacidade máxima de 20 clientes, população infinita e atendimento randômico. A notação condensada **A/B/c** supõe que a capacidade máxima do sistema e a população são infinitas, e que a disciplina da fila é FIFO.

Os principais modelos de filas, que serão apresentados, são os modelos de Poisson (M/M/1 ou M/M/c) e o modelo Erlang.

A.7. O MODELO M/M/1

O modelo M/M/1 é um dos mais simples. Supõe-se que os ritmos de chegada e de atendimento são Markovianos, ou seja, que seguem a distribuição de Poisson ou a exponencial negativa, com um único atendente. Consideraremos os casos em que temos a população infinita e finita.

A figura abaixo mostra uma representação da fila para o modelo M/M/1. O retângulo tracejado representa o sistema, e o quadrado representa o cliente sendo atendido.

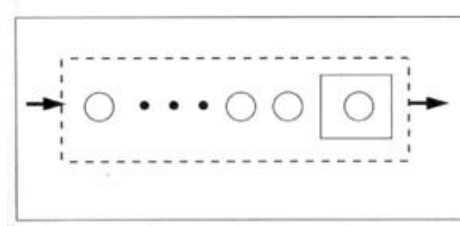



Figura A-7.1: Representação do modelo M/M/1

A.7.1. O Modelo M/M/1 com População Infinita

Inicialmente, consideramos que a população que fornece clientes é infinita. Neste caso, temos as seguintes relações:

- **Número médio de clientes na fila:**

$$\text{NF} = \frac{\lambda^2}{\mu(\mu - \lambda)} \qquad \text{(A.25)}$$

- **Número médio de clientes no sistema:**

$$\text{NS} = \frac{\lambda}{\mu - \lambda} \qquad \text{(A.26)}$$

- **Tempo médio que o cliente fica na fila:**

$$\text{T}F = \frac{\lambda}{\mu(\mu - \lambda)} \qquad \text{(A.27)}$$

- **Tempo médio que o cliente fica no sistema:**

$$\mathrm{T}S = \frac{1}{\mu - \lambda} \tag{A.28}$$

- **Probabilidade de existirem n clientes no sistema:**

$$P_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n \tag{A.29}$$

- **Taxa de utilização:**

$$\rho = \frac{\lambda}{\mu} \tag{A.30}$$

Como já vimos, em sistemas estáveis temos $\mu > \lambda$, ou $\rho < 1$. Quando ρ tende a 1, a fila tende a aumentar infinitamente, como mostra a seguinte relação:

$$\mathrm{NF} = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho} \tag{A.31}$$

A figura a seguir mostra a relação entre NF e ρ, onde percebemos que o crescimento de NF é exponencial, o que implica em crescimento nos tempos do cliente na fila (TF) e no sistema (TS). É importante observar que, se o sistema está saturado (ρ próximo de 1), é possível reduzir a fila dobrando a capacidade de atendimento, o que faz com que ρ, neste caso, fique menor que 0,5.

Figura A-7.2: Relação entre NF e ρ para o modelo M/M/1

Como exemplo deste modelo, suponhamos que as chegadas a uma cabine telefônica obedeçam à lei de Poisson, com ritmo de 6 chegadas por hora. A duração média de um telefonema é de 3 minutos, e suponhamos que siga a distribuição exponencial. Como $\lambda = 6$ chegadas/hora e TA = 3 minutos, obtemos IC (intervalo entre chegadas) = 10 minutos e μ = 20 atendimentos/hora. Calculando NF, NS e TF, obtemos:

$$\text{NF} = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{6^2}{20(20-6)} = \frac{36}{20 \cdot 14} = 0{,}128 \text{ pessoas na fila}$$

$$\text{NS} = \frac{\lambda^2}{\mu-\lambda} = \frac{6^2}{20-6} = \frac{36}{14} = 0{,}428 \text{ pessoas no sistema}$$

$$\text{TF} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{6}{20(20-6)} = \frac{6}{20 \cdot 14} = 0{,}021 \text{ horas na fila} = 1{,}28 \text{ minutos}$$

A.7.2. O Modelo M/M/1/K com População Finita

Um caso particular do modelo M/M/1 é aquele em que a população que fornece clientes é considerada finita. Assim, a notação do modelo torna-se M/M/1/K, sendo K o tamanho da população.

Com a população considerada finita, as relações obtidas para o caso anterior, com população infinita, tornam-se as seguintes, sendo P_0 a probabilidade de um cliente ser atendido imediatamente, sem ter que entrar em fila:

- **Número médio de clientes na fila:**

$$\text{NF} = K - \frac{\lambda+\mu}{\lambda} + (1-P_0) + \frac{\lambda}{\mu} \quad \text{(A.32)}$$

- **Número médio de clientes no sistema:**

$$\text{NS} = K - \frac{\lambda+\mu}{\lambda} + (1-P_0) + \frac{\lambda}{\mu} + \frac{\lambda}{\mu} \quad \text{(A.33)}$$

- **Tempo médio que o cliente fica na fila:**

$$\text{T}F = \frac{K}{\lambda} - \frac{(\lambda + \mu)(1 - P_0)}{\lambda^2} \tag{A.34}$$

- **Tempo médio que o cliente fica no sistema:**

$$\text{T}F = \frac{K}{\lambda} - \frac{(\lambda + \mu)(1 - P_0)}{\lambda^2} + \frac{1}{\mu} \tag{A.35}$$

- **Probabilidade de existirem n clientes no sistema:**

$$P_n = \frac{\left(\frac{\mu}{\lambda}\right)^{K-n}}{(K-n) \cdot \sum_{j=0}^{K} \frac{\left(\frac{\mu}{\lambda}\right)^j}{j!}} \tag{A.36}$$

- **Taxa de utilização:**

$$\rho = \frac{\lambda}{\mu} \tag{A.37}$$

A.8. O MODELO M/M/c

Este modelo de fila é bastante semelhante ao modelo M/M/1. Os ritmos de chegada e atendimento também seguem a distribuição de Poisson ou a distribuição exponencial negativa. A diferença é que, neste modelo, temos c atendentes. A figura abaixo representa o modelo M/M/c, onde vemos que há uma única fila. Supondo que a capacidade de atendimento de cada atendente é a mesma, apresentaremos os casos em que consideramos a população finita e infinita.

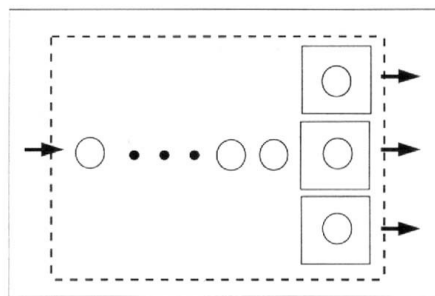



Figura A-8.1: Representação do modelo M/M/c

A.8.1. O Modelo M/M/c com População Infinita

As fórmulas para o modelo M/M/c são bastante complexas e difíceis de serem manipuladas. Desta forma, torna-se mais simples examinar este modelo através de gráficos.

No modelo M/M/c, temos c servidores. Assim, a taxa de utilização é:

$$\rho = \frac{\lambda}{c \cdot \mu} \tag{A.38}$$

As Figuras A-8.2 e A-8.3 mostram os gráficos do número médio de clientes na fila (NF) e do número de clientes no sistema (NS), respectivamente, ambos em função de ρ , variando-se o número de servidores c.

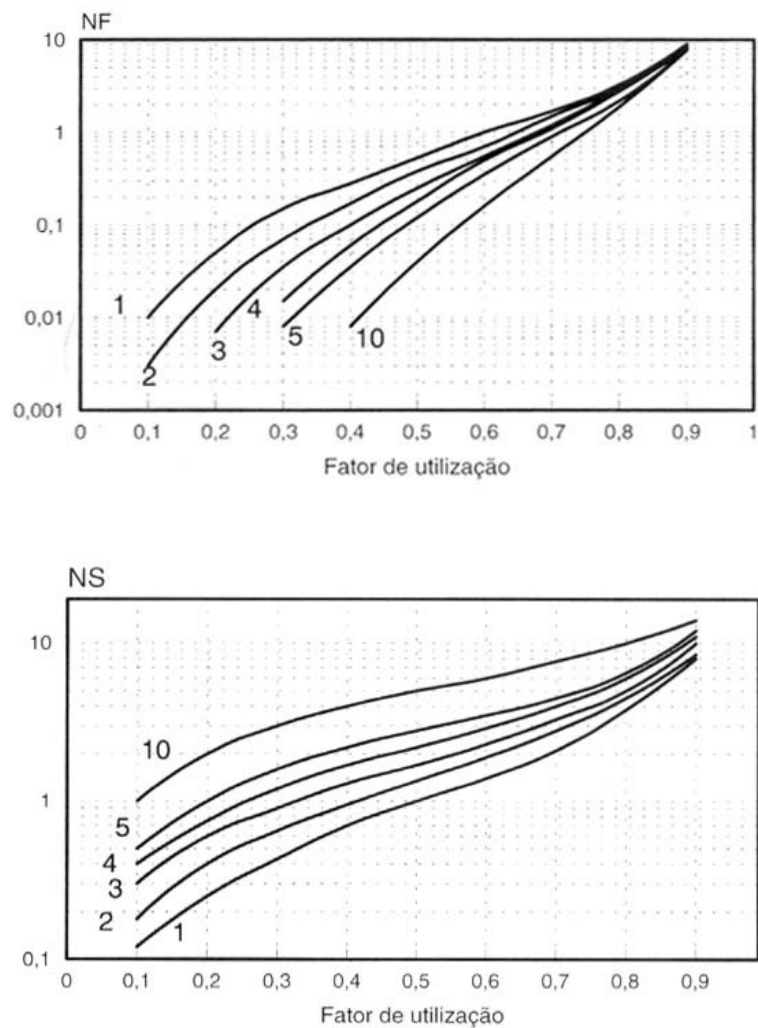



Figura A-8.3: Relação entre NS e ρ para o modelo M/M/1

Em ambos os gráficos, a ordenada, mostrada em escala logarítmica, tende a infinito quando ρ tende a 1, como no caso do modelo M/M/1. Também observamos a grande redução no número de clientes quando dobramos a capacidade de atendimento, ou seja, reduzimos ρ pela metade.

Como exemplo deste modelo, consideremos o caso de um banco que deseja modificar a forma de atendimento de seus clientes, que hoje funciona com diversas filas, pela introdução do sistema de fila única. Os dados atuais são: λ = 70 clientes/hora, distribuídos em 5 filas; c=5 atendentes; μ = 20 clientes/hora. Assim, na situação atual, temos:

$$\lambda = \frac{70}{5} = 14$$

$$TS = \frac{1}{\mu - \lambda} = 0{,}167 \text{ hora} = 10 \text{ minutos}$$

$NS = \lambda \cdot TS = 14 \cdot 0{,}167 = 2{,}33$ pessoas

$NS_{total} = 5 \cdot NS = 5 \cdot 2{,}33 = 11{,}67$ pessoas

Na situação com fila única, teríamos:

$$\rho = \frac{\lambda}{c \cdot \mu} = \frac{70}{5 \cdot 20} = 0{,}7$$

$NS = 5$ pessoas (a partir da Figura A-4.4)

$TS = \frac{NS}{\lambda} = \frac{5}{70} = 0{,}07$ hora = 4,3 minutos

Assim, vemos que um sistema com fila única é mais eficiente que um sistema com diversas filas, pois temos, na média, um menor número de clientes no sistema e menor tempo de atendimento por cliente. Esta conclusão foi aplicada há pouco tempo (este texto está sendo escrito no final de 2003) em agências bancárias e nos Correios. Os supermercados não o fazem por causa da distribuição física das caixas (em geral numa fila paralela à porta de entrada) que facilita o estabelecimento de várias filas que, espera-se, sejam curtas.

A.8.2. O Modelo M/M/c/K com População Finita

No caso em que temos a população finita, há algumas diferenças em comparação com o caso em que consideramos a população infinita. Para efeitos práticos, quando o tamanho da população alcança determinado valor, o número de clientes na fila torna-se o mesmo do caso com população infinita.

Tendo em vista a complexidade matemática deste modelo, não iremos apresentar detalhes dos cálculos.

A.9. O MODELO ERLANG

No modelo Erlang, o ritmo de chegada dos clientes segue a distribuição de Poisson, como nos modelos anteriores, mas o atendimento segue a distribuição Erlang de grau m. O modelo será comparado com os outros, através do dimensionamento de equipamentos.

O dimensionamento de equipamentos deve buscar a minimização do tempo de espera na fila e do custo do sistema e a maximização da capacidade de produção. Assim, o processo de dimensionamento deve procurar a quantidade ideal de atendentes que produza um valor adequado para TF. Veremos que, para um dado valor de TF, o modelo M/Em/c necessita de menos atendentes que o modelo M/M/c, ou seja, este modelo superdimensiona os equipamentos em relação ao modelo Erlang.

A.9.1. O Modelo M/Em/1

Nesta primeira situação, supomos o modelo Erlang com apenas um atendente. A figura abaixo mostra o gráfico da distribuição de Erlang. Observamos que, quando m=1, esta distribuição tem o mesmo formato da exponencial negativa, e, conforme m cresce, a distribuição tende para a normal. Se m tende a infinito, a função tende a uma constante (TA), ou seja, quanto maior o valor de m, mais constante se torna o tempo de atendimento. Assim, m pode ser visto como um medidor da ordem/desordem do tempo de atendimento.

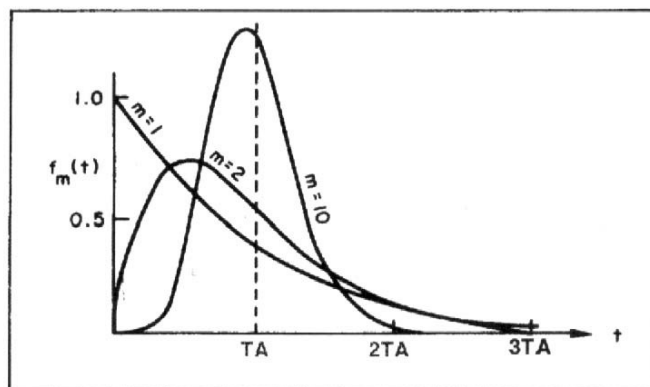



Figura A-9.1: Gráfico da distribuição de Erlang

A Figura A-9.2 mostra o tamanho da fila (NS) em função da taxa de utilização ρ para as distribuições exponenciais, Erlang-2, Erlang-5 e constante. Podemos observar que, para valores de ρ abaixo de 0,3, todas as curvas são praticamente iguais. Para ρ maior que 0,3, a

diferença entre as curvas torna-se significativa, sendo que a distribuição exponencial obtém valores de NS – e, conseqüentemente, de NF, TF e TS – maiores que a distribuição Erlang.

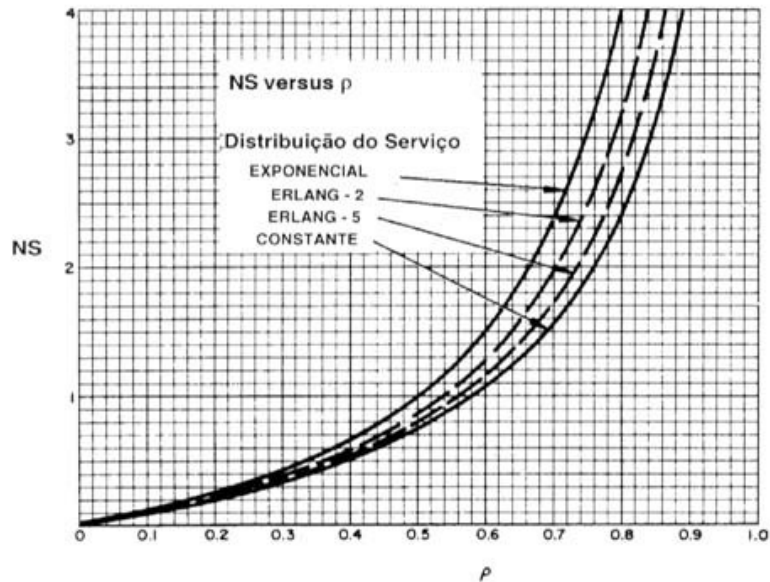



Figura A-9.2: Relação entre NS e ρ para o modelo M/Em/1

Na Figura A-9.3, vemos TS/TA em função de ρ para as distribuições hiperexponencial, exponencial, Erlang-2, Erlang-5 e constante. Como os gráficos são semelhantes aos da Figura A.18, as conclusões são as mesmas que foram obtidas para esta.

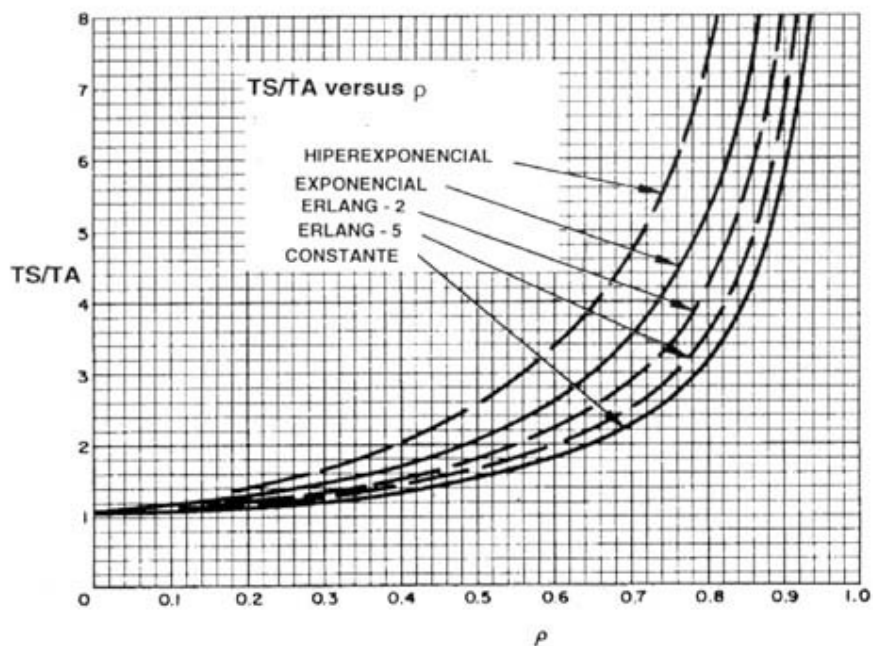



Figura A-9.3: Relação entre TS/TA e ρ para o modelo M/Em/1

Supondo, como exemplo, um sistema com $\lambda = 2$ clientes/minuto e $\mu = 3$ clientes/minuto, temos, para cada modelo de fila, os seguintes valores para NS, obtidos a partir da Figura A-9.2:

Distribuição	NS
Exponencial	2,10
Erlang-2	1,80
Erlang-5	1,55
Constante	1,40

A.9.2. O Modelo M/Em/c

O modelo anterior considerava o sistema com apenas um atendente. Se generalizarmos para o caso em que tivermos c atendentes, obtemos resultados semelhantes aos do modelo M/M/c. As Figuras A-9.4 e A-9.5 mostram, respectivamente, NF e NS em função de ρ .

As semelhanças entre os modelos podem ser comprovadas comparando-se as Figuras A-9.4 e A-9.5 com as Figuras A-8.2 e A-8.3.

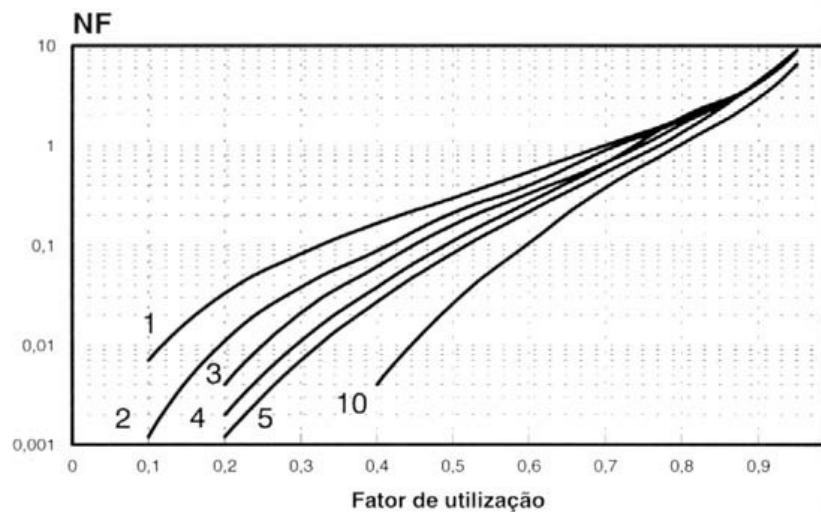
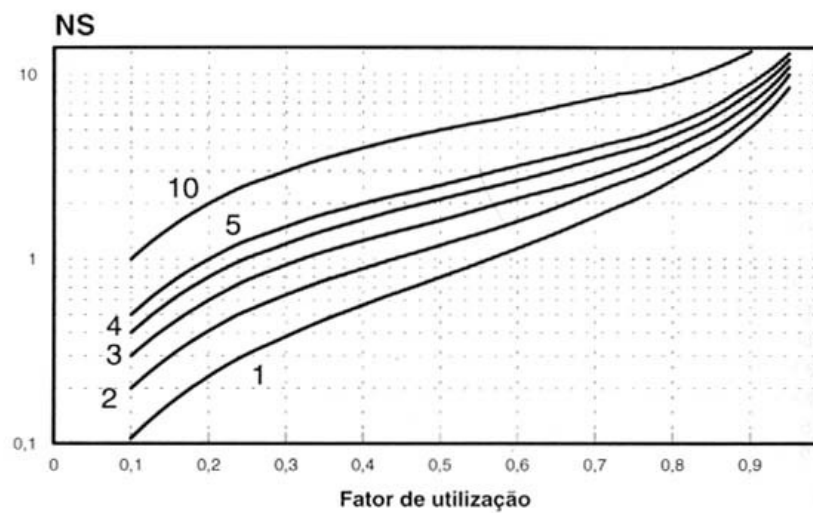



Figura A-9.4: Relação entre NF e ρ para o modelo M/Em/c

Como exemplo, temos o sistema de filas seqüenciais em uma fábrica, mostrado na Figura A-9.6. O sistema possui três servidores, nas etapas de fabricação, inspeção e reparo de equipamentos.

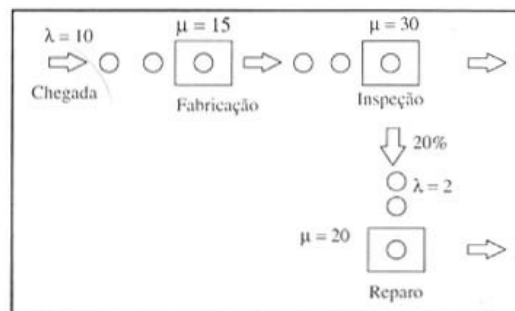



Figura A-9.6: Seqüência de filas em uma fábrica

A partir da Figura A-9.4, as filas formadas em cada servidor do sistema são:

Fila	ρ	NF
Produção	0,66	0,85
Inspeção	0,33	0,11
Reparo	0,10	0,007

Apêndice B

COMUNICAÇÃO DIGITAL

B.1. MODULAÇÃO E TRANSMISSÃO DIGITAL

Na transmissão digital por um canal passa-banda a seqüência de bits que se deseja transmitir é modulada em uma portadora, geralmente senoidal. com limites de freqüência impostos pelo canal.

A informação binária é transmitida mudando-se a fase, a freqüência ou a amplitude da portadora. Estes três esquemas básicos são conhecidos como phase-shift keying, frequency-shift keying e amplitude-shift keying, respectivamente.

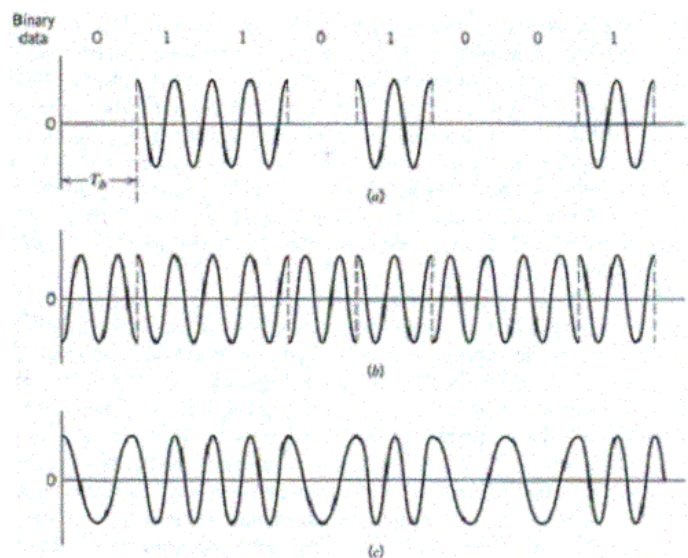



Figura B-1.1: (a) amplitude-shift keying (b) phase-shift keying (c) frequency-shift keying

Algumas medidas são usadas para comparar estas três formas de transmissão. São elas: a probabilidade de erro, a densidade espectral de potência e a eficiência de banda.

O maior objetivo no projeto dos sistemas de transmissão digital é minimizar a probabilidade de erro, que consiste no fato do emissor enviar uma mensagem, dentro de um conjunto de mensagens, que é modificada pelo ruído do canal, usualmente modelado como ruído branco aditivo gaussiano (AWGN) de tal forma que o receptor decida por uma outra

mensagem deste mesmo conjunto de mensagens. A probabilidade varia inversamente com a potência de transmissão e diretamente com o nível de ruído no canal, mas também depende da forma de modulação utilizada. Como não há controle sobre o nível de ruído do canal e como a potência de transmissão está relacionada com a energia e, consequentemente, com custo o estudo das formas de transmissão se torna importante.

A outra condição importante na transmissão digital é o uso racional da banda. Quanto menos banda uma transmissão utilizar, a faixa disponível de freqüências poderá ser dividida por um número maior de canais permitindo que mais sistemas estejam conectados ao mesmo tempo. A densidade espectral de potência se constitui em uma função energia X banda. Com ela podemos visualizar como a banda está sendo utilizada na transmissão. Quanto mais energia estiver concentrada no trecho considerado da banda mais eficiente será a transmissão.

Neste mesmo cenário entra a medida da eficiência de banda. Esta medida, mais simples que a densidade espectral de potência, é definida como a razão entre a taxa de bits (número de bits transmitidos por unidade de tempo) R_b e a banda utilizada B.

$$\rho = \frac{R_b}{B} \quad bits/s/Hz$$

Dos três esquemas básicos de transmissão que utilizam portadora, a modulação em fase é a que geralmente possui a melhor relação entre probabilidade de erro pequena e eficiência no uso da banda, sendo ela e suas derivações as formas de transmissão mais utilizadas pelos sistemas digitais.

Neste Apêndice apresentaremos a modulação em fase (PSK) e algumas de suas derivações. Não entraremos nos detalhes de cálculo da probabilidade de erro e da densidade espectral de potência. Apenas citaremos seus valores como figuras de mérito para a comparação entre os sistemas apresentados.

B.2. PHASE SHIFT-KEYING BINÁRIO E COERENTE (PSK)

Definiremos o par de sinais $s_1(t)$ e $s_2(t)$ para representar os símbolos binários 0 e 1 respectivamente, definidos por:

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$
$$s_2(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$

Onde f_c é a freqüência da portadora, E_b é a energia por bit e T_b é o período de transmissão de cada mensagem (bit), que deve ser um múltiplo do período $\frac{1}{f_c}$.

A raiz quadrada que multiplica o cosseno é apenas para normalizar a energia do símbolo.

Notemos que a diferenciação entre os símbolos '0' e '1' é feita apenas mudando a fase do cosseno em π radianos. Dentro do período de transmissão Se o receptor receber um sinal em fase com o seu oscilador, de mesma freqüência f_c, ele deve optar pelo símbolo '0'. Se ele receber um sinal em contra-fase com o seu oscilador ele deve optar pelo símbolo '1'.

Neste ponto tempos um aspecto importante do PSK, e também do ASK e do FSK: O receptor precisa estar sincronizado com o a portadora gerada pelo emissor, o que complica o *design* do receptor, requerendo o uso de PLL's, por exemplo. Por isso o termo coerente no título deste item. Mais à frente, mostraremos uma derivação do PSK onde este sincronismo não é necessário, ou seja, a recepção é não-coerente.

O espaço de estados do PSK binário é constituído de uma base, ou seja , é um espaço de uma dimensão (reta). Esta base é definida por:

$$\phi(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t)$$

e os pontos no espaço de estados dos símbolos $s_1(t)$ e $s_2(t)$ são dados por:

$$'0' \longrightarrow s_1(t) = \sqrt{E_b}\,\phi(t)$$
$$'1' \longrightarrow s_2(t) = -\sqrt{E_b}\,\phi(t)$$

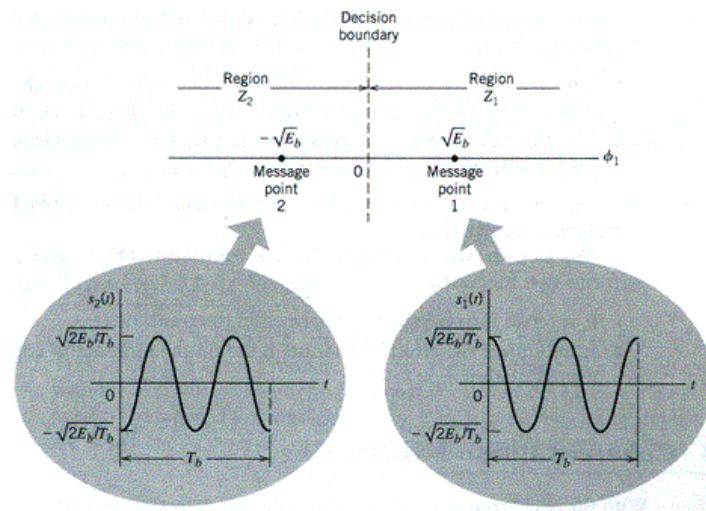



Figura B-2.1: Pontos de mensagem $s_1(t)$ e $s_2(t)$ no espaço de estados do PSK binário

A decisão do receptor, pela ótica do espaço de estados, é feita passando-se o sinal recebido dentro de um período T_b por um filtro casado com a base $\phi(t)$. Se o resultado for positivo decide-se pela mensagem $s_1(t)$, caso seja negativo decide-se pela mensagem $s_2(t)$

A probabilidade de erro é dada por $P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right)$ onde $erfc$ é a função gaussiana de média zero e variância 1 e N_0 é uma medida do nível de ruído do canal.

A densidade espectral de potência é dada por $S_B(f) = 2E_b sinc^2(T_b f)$.

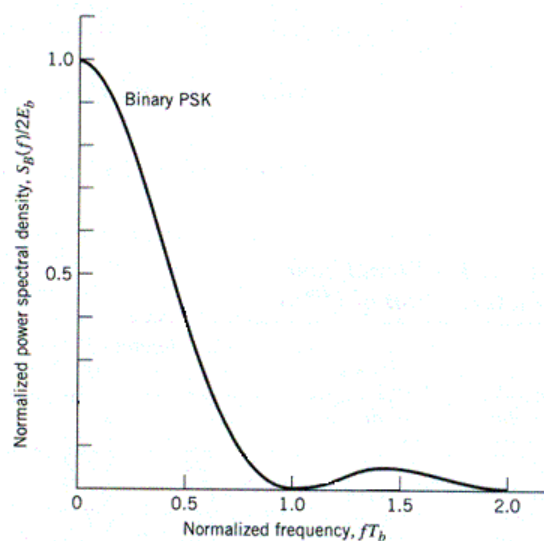



Figura B-2.2: Densidade espectral de potência do PSK binário

Notemos que a maior parte da energia está concentrada no nódulo principal. Quanto mais estreito for o nódulo, maior será a eficiência na utilização da banda.

A modulação em fase, desta forma, se confunde com a modulação em amplitude (ASK) pois mudar a fase da portadora em π radianos equivale a multiplicar a amplitude da portadora por –1. Portanto, tanto podemos considerar que a informação está contida na fase como podemos considerar que a informação está contida na amplitude da portadora.

B.3. QUADRATURE PHASE-SHIFT KEYING, COERENTE (QPSK)

No PSK binário o transmissor utiliza apenas dois valores possíveis para a fase da senóide transmitida: 0 e π radianos. O Quadrature phase-shift keying (QPSK) utiliza quatro valores para a fase. São eles: $\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$ e $\frac{7\pi}{4}$. Desta forma transmitimos não apenas um, mais dois bits (dibit) por cada período T de transmissão de cada mensagem.

Definimos os sinais transmitidos por:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right] \qquad i = 1,2,3,4$$

Onde f_c é a freqüência da portadora, E é a energia por mensagem (dibit) e $T = \frac{1}{f_c}$ é o período de transmissão de cada mensagem.

Para o estudo do espaço de estados utilizaremos uma relação trigonométrica. Sabemos que:

$$\cos(a+b) = \cos(a)\cdot\cos(b) - \operatorname{sen}(a)\cdot\operatorname{sen}(b)$$

Aplicando esta relação sobre o sinal transmitido $s_i(t)$ temos:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right] =$$
$$\sqrt{\frac{2E}{T}} \cos\left[(2i-1)\frac{\pi}{4}\right]\cos(2\pi f_c t) - \sqrt{\frac{2E}{T}} \operatorname{sen}\left[(2i-1)\frac{\pi}{4}\right]\operatorname{sen}(2\pi f_c t)$$

Olhando para a expressão acima definimos duas bases que constituirão o espaço de estados do QPSK, ou seja, temos um espaço de duas dimensões (plano). Estas bases são definidas por:

$$\begin{cases} \phi_1(t) = \sqrt{\dfrac{2}{T}}\cos(2\pi f_c t) \\ \phi_2(t) = \sqrt{\dfrac{2}{T}}\,\text{sen}(2\pi f_c t) \end{cases}$$

Notemos que $\phi_1(t)$ e $\phi_2(t)$ constituem um conjunto de bases pois elas são ortogonais, ou seja, a autocorrelação entre elas é zero:

$$\int_t^{t+T} \phi_1(t)\cdot\phi_2(t)\cdot dt = 0$$

Prova:

$$\int_t^{t+T} \phi_1(t)\cdot\phi_2(t)\cdot dt = \frac{2}{T}\int_t^{t+T}\cos(2\pi f_c t)\cdot\text{sen}(2\pi f_c t)\cdot dt = \frac{1}{T}\int_t^{t+T}\text{sen}(4\pi f_c t)\cdot dt = 0$$

pois a integral do seno sobre um número inteiro de períodos vale zero

Desta forma expressamos o sinal QPSK através da representação de estados:

$$s_i(t) = \sqrt{E}\cos\left[(2i-1)\frac{\pi}{4}\right]\phi_1 - \sqrt{E}\,\text{sen}\left[(2i-1)\frac{\pi}{4}\right]\phi_2$$

Os pontos no espaço de estados dos símbolos $s_i(t)$ ($i = 1,2,3,4$) são dados por:

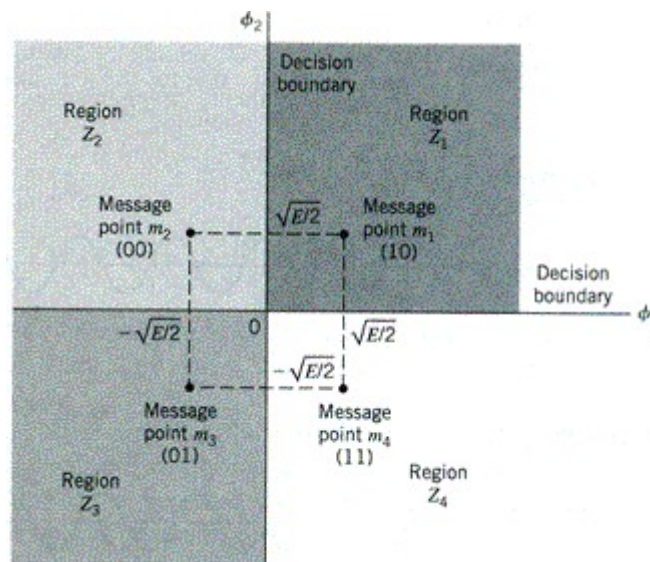



Figura B-3.1: Pontos de mensagem $s_i(t)$ no espaço de estados do QPSK

Estabelecemos a relação entre os pontos de mensagem e os valores do dibit utilizando o código de Gray:

Dibit (código de gray)	Fase do sinal QPSK	Coordenadas no plano de estados s_1	s_2
10	$\frac{\pi}{4}$	$\sqrt{\frac{E}{2}}$	$-\sqrt{\frac{E}{2}}$
00	$\frac{3\pi}{4}$	$-\sqrt{\frac{E}{2}}$	$-\sqrt{\frac{E}{2}}$
01	$\frac{5\pi}{4}$	$-\sqrt{\frac{E}{2}}$	$\sqrt{\frac{E}{2}}$
11	$\frac{7\pi}{4}$	$\sqrt{\frac{E}{2}}$	$\sqrt{\frac{E}{2}}$

A probabilidade de erro é aproximadamente $P_e \cong \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right)$. Notemos que a probabilidade de erro é a mesma do PSK binário, esta é uma grande vantagem do QPSK. Utilizando o QPSK ao invés do PSK binário dobramos a taxa de transferência (pois estamos transmitindo dois bits ao invés de um em um mesmo período de tempo) sem que, para isso, aumentemos a probabilidade de errar um desses bits. Desta forma dobramos a eficiência de banda da transmissão.

A densidade espectral de potência é dada por $S_B(f) = 4E_b sinc^2(2T_b f)$

B.3.1. $\frac{\pi}{4}$-Shifted QPSK

Chegou a hora de conhecermos uma transmissão não coerente, ou seja, onde o receptor não precisa estar sincronizado com o transmissor. O $\frac{\pi}{4}$-Shifted QPSK é um QPSK onde o transmissor, ao transmitir um novo dibit, alterna entre duas constelações de pontos de mensagens diferentes. São elas:

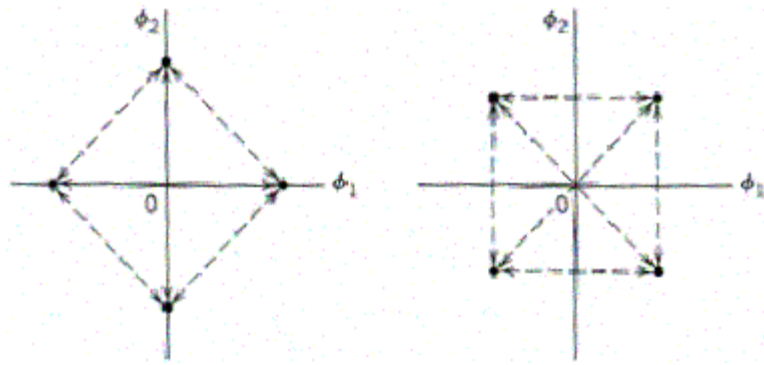



Figura B-3.2: As duas constelações do $\frac{\pi}{4}$-Shifted QPSK

O transmissor muda a fase da senóide em $\pm\frac{\pi}{4}$ ou $\pm\frac{3\pi}{4}$ sempre que um novo dibit é enviado, segundo a tabela abaixo:

dibit	Mudança na fase ($\Delta\theta$)
10	$\frac{\pi}{4}$
00	$\frac{3\pi}{4}$
01	$\frac{-3\pi}{4}$
11	$\frac{-\pi}{4}$

O receptor, ao detectar uma mudança de fase no sinal recebido, decide pela nova mensagem recebida em função do quanto a fase mudou. Desta forma o receptor não precisa estar sincronizado com o transmissor diminuindo muito a sua complexidade.

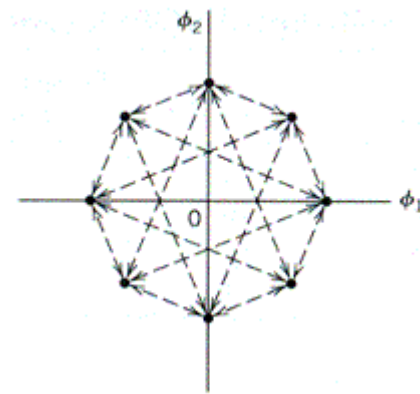



Figura B-3.3: Possíveis transições do $\frac{\pi}{4}$-Shifted QPSK

B.4. PSK M-ÁRIO

Estudaremos a modulação por fase para um número M de pontos de mensagens. Definimos o sinal transmitido para cada um desses M pontos como sendo:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + \frac{2\pi}{M}(i-1)\right] \text{ com } \quad i = 1,2 \ldots M$$

O estudo do espaço de estados é análogo ao que foi feito no QPSK. De fato o QPSK é um caso particular do PSK M-ário onde M = 4. As bases são dadas por:

$$\begin{cases} \phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t) \\ \phi_2(t) = \sqrt{\frac{2}{T}} \operatorname{sen}(2\pi f_c t) \end{cases}$$

e o sinal transmitido é dado por :

$$s_i(t) = \sqrt{E} \cos\left[\frac{2\pi}{M}(i-1)\right] \cdot \phi_1(t) - \sqrt{E} \operatorname{sen}\left[\frac{2\pi}{M}(i-1)\right] \cdot \phi_2(t) \qquad \text{com} \quad i = 1,2 \ldots M$$

Todos os M pontos de mensagem estão contidos em um círculo de raio $\sqrt{E}$ (todos os pontos de mensagem têm a mesma energia). O ângulo do ponto de mensagem, representado em coordenadas polares, corresponde à fase da senóide transmitida para aquele ponto. Como exemplo é mostrada abaixo a constelação do PSK M-ário para M = 8.

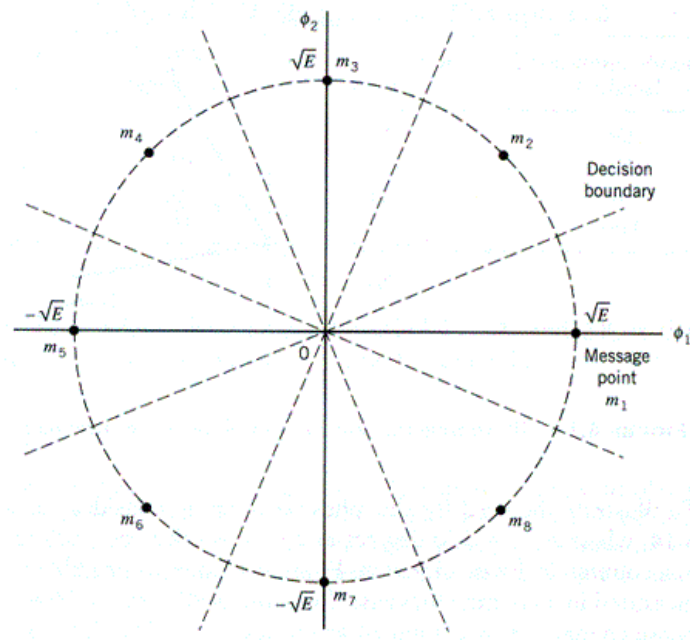



Figura B-4.1: Constelação do PSK M-ário para M = 8

Pela figura acima percebemos que, quanto maior o número M de pontos, mais próximos estes pontos estarão uns dos outros e, consequentemente, maior será a chance do receptor decidir por uma mensagem errada na presença de ruído. Dividindo-se o plano em M fatias (marcando o limite de decisão do receptor) vemos que o ruído pode alterar o ângulo da senóide transmitida em, no máximo, $\frac{\pi}{M}$ radianos para que o receptor não erre em sua decisão. Abaixo é mostrada a área de decisão de um ponto de mensagem para o PSK M-ário com M = 8.

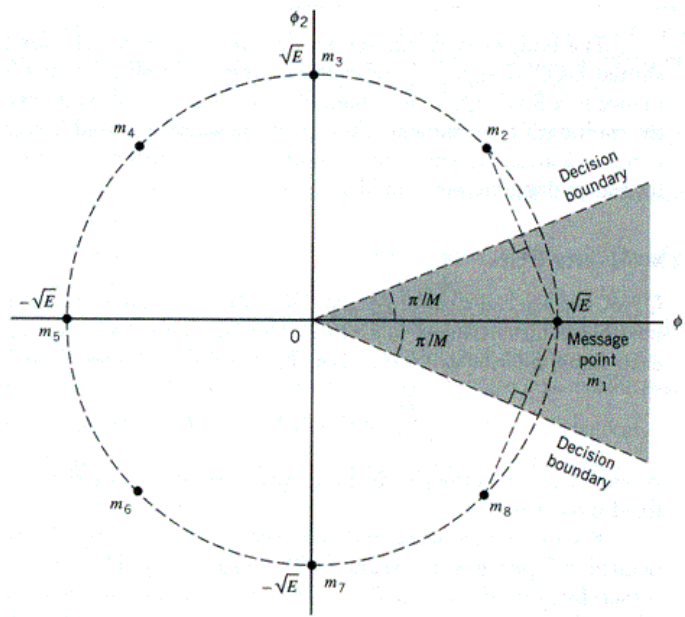



Figura B-4.2: Área correspondente a decisão pela mensagem m_1 para o PSK M-ário com M = 8

A probabilidade de erro é aproximadamente:

$$P_e \cong erfc\left[\sqrt{\frac{E}{N_0}}\,\text{sen}\left(\frac{\pi}{M}\right)\right]$$

e a densidade espectral de potência é dada por :

$$S_B(f) = 2E_b \cdot \log_2 M \cdot sinc^2\left(T_b \cdot f \cdot \log_2 M\right)$$

ambas são função de M, como não poderia deixar de ser! Abaixo é mostrada a densidade espectral de potência para M = 2 (PSK binário), M = 4 (QPSK) e M=8.

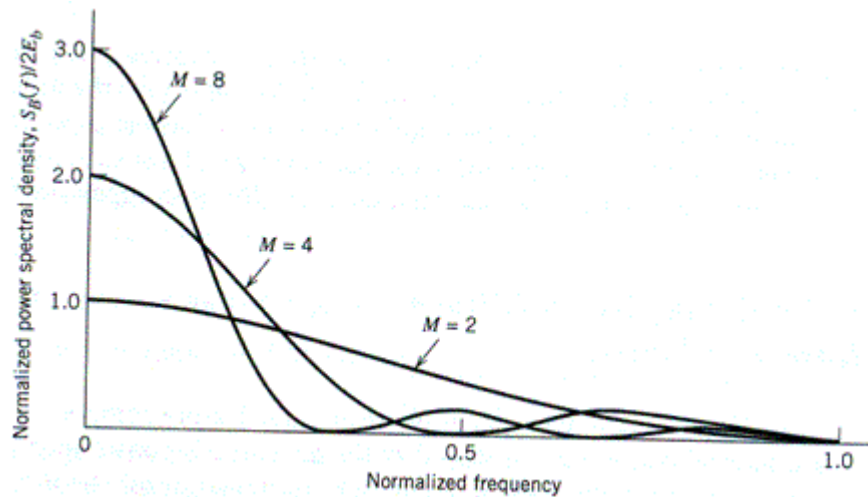



Figura B-4.3: Densidade espectral de potência para o PSK M-ário

Quanto maior for o número M de pontos, mais concentrada estará a energia em função da banda, aumentando a eficiência, no entanto, maior será a probabilidade de erro.

B.5. QUADRATURE AMPLITUDE MODULATION (QAM)

Em todas as modulações apresentadas mantivemos a amplitude do sinal transmitido constante. Apresentaremos agora uma modulação híbrida onde a informação está contida tanto na fase quanto na amplitude da senóide.

Mantendo as bases usadas no QPSK e no PSK M-ário:

$$\begin{cases} \phi_1(t) = \sqrt{\dfrac{2}{T}}\cos(2\pi f_c t) \\ \phi_2(t) = \sqrt{\dfrac{2}{T}}\,\mathrm{sen}(2\pi f_c t) \end{cases}$$

definimos o sinal QAM como sendo:

$$s_i(t) = \sqrt{E_0}\cdot a_i \cdot \phi_1(t) - \sqrt{E_0}\cdot b_i \cdot \phi_2(t)$$

onde E_0 é a energia da mensagem com menor energia (ponto mais próximo da origem).

a_i e b_i são inteiros e definem a posição do ponto da mensagem no espaço de estados (plano). A fim de se minimizar a probabilidade de erro, distribuem-se os pontos de mensagem igualmente espaçados em um quadrado e centraliza-se este quadrado na origem com o fim de minimizar a energia média da constelação. Abaixo é mostrada uma distribuição QAM com M = 16, conhecida como QAM-16.

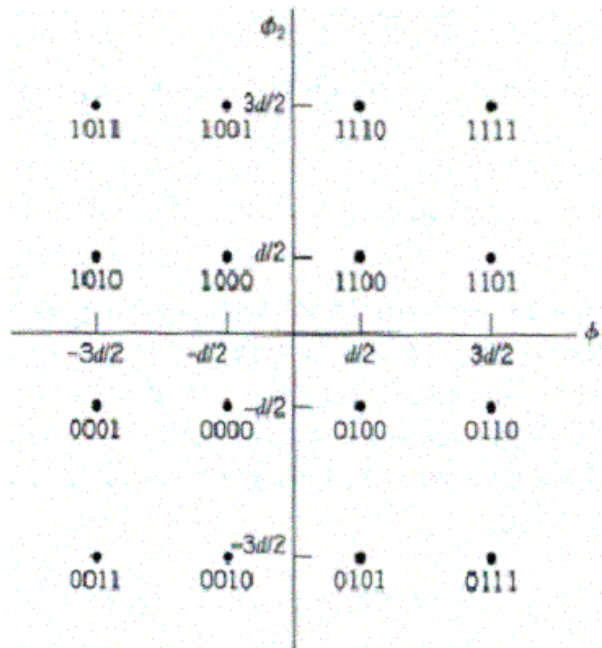



Figura B-5.1: Constelação do QAM-16

Repare que se utiliza o código de Gray, em cada quadrante, para os dois bits menos significativos e que os dois bits mais significativos são iguais em cada quadrante; e que estes bits também estão codificados pelo código de Gray em relação aos 4 quadrantes. Esta codificação minimiza a probabilidade de erro.

Os QAM com M=16 e M=32 são muito usados por modems de computador que transmitem dados pela linha telefônica (padrões V.16 e V.32)

Os pontos de mensagem não possuem a mesma energia pois não possuem a mesma distância até a origem (não estão distribuídos em uma circunferência como no PSK M-ário). Surge a necessidade de se definir a energia média da constelação como sendo:

$$E_{AV} = \frac{2(M-1)E_0}{3}$$

então define-se a probabilidade de erro em função da energia média:

$$P_e \cong 2 \cdot \left(1 - \frac{1}{\sqrt{M}}\right) \cdot erfc\left[\sqrt{\frac{3E_{AV}}{2(M-1)N_0}}\right]$$

Abaixo são mostradas as constelações do PSK M-ário e do QAM M-ário para M = 16.

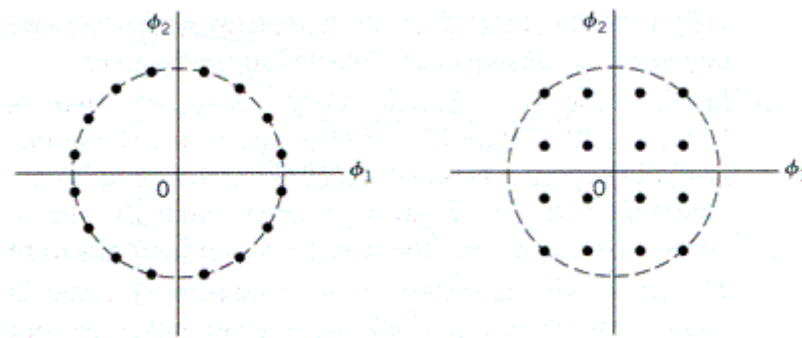



Figura B-5.2: Constelações do PSK-16 e do QAM-16

Bibliografia

[1] ALENCAR, Marcelo Sampaio de. *Telefonia Digital*. 4ª ed. São Paulo: Editora Érica, 1998.

[2] ALENCAR, Marcelo Sampaio de. *Sistemas de Comunicações*. 1ª ed. Editora Érica, 2001.

[3] RUIZ, Roberto. *O Telefone, uma das mais simples e fabulosas criações do engenho humano*. Celta Ed. e Publicidade, CTB, 1973.

[4] HALL, Donald B. *Basic Acoustics*. California State University: Harper & Row Publishers.

[5] DINIZ, Paulo S. R. *Principles of Wireless Communications*. Apostila, 2003.

[6] REIMEIRS, Ulrich. *Digital Video Broadcasting: The International Standard for Digital Television*. 1ª ed. Springer Verlag, 2001.

[7] HAYKIN, Simon. *Communication Systems*. 4ª ed. New York: John Wiley & Sons, 2001.

[8] TANENBAUM, Andrew S. *Computer Networks*, 4ª edição. Upper Saddle River, NJ. Prentice Hall, 2002.

[9] FERRARI, Antônio. *Telecomunicações. Evolução e Revolução*. Editora Érica.

[10] SCHLZRINNE, H., CASNER, S., FREDERICK, R., E JACOBSON, V. *RTP: A transport protocol for real-time applications*. Internet RFC 1889, janeiro de 1996.

[11] SOUZA, L. B. de. *Redes CISCO CCNA (CISCO Certified Netork Associate)*. Ed. Érica, 2002.

[12] HERSENT, O., GURLE, D., E PETIT, J.-P. *IP Telephony - Packet-based multimedia communications systems*, first ed. Addison-Wesley Publishing Company, 1997.

[13] ROSENBORG, J., SCHULZRINNE, H., CAMARILLO G., JOHNSTON A., PETERSON, J., SPARKS, R., HANDLEY, M., E SCHOOLER, E. *SIP. Session Initiation Protocol*. Internet RFC 3261, junho de 2002.

[14] ITU-T. G.711: Pulse code modulation (PCM) of voice frequencies. In *International Telecommuication Union*, novembro de 1988.

[15] ITU-T. G.723.1: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. In *International Telecommuication Union*, março de 1996.

[16] ITU-T. G.729: Coding of speech at 8kbit/s using conjugate-structure algebraic code excited linear prediction. In *International Telecommuication Union*, março de 1996.

[17] ITU-T. One-way transmission time. In *Recomendação* G.114, março de 1993.

[18] KAO, Charles K. *Optical Fiber Systems*. McGraw-Hill Book Company, 1988.

[19] GIOZZA, William F., CONFORTI, Evando, WALDMAN Hélio. *Fibras Ópticas: Tecnologia e Projeto de Sistemas*. McGraw-Hill Book Company, 1991.

[20] SILVA, Loreno, BRANDÃO, Waldemir, NOBRE, Celso. *Sistema de sinalização por canal comum CCITT nº 7*" – Telebrás 1983.

[21] BELLAMY, John. *Digital Telephony*. 3a ed., Wiley-Interscience, 2000.

[22] GARCIA, Alberto Leon. *Probability and Random Processes for Electrical Engineering*. 2ª ed. New York: Addison-Wesley, 1994.

[23] PRADO, Darcy Santos do. *Teoria das Filas e da Simulação*. Belo Horizonte: Editora de Desenvolvimento Gerencial, 1999.

[24] ALLES, Antony. ATM Internetworking. Cisco Systems, Inc. May, 1995. Disponível em julho de 1998 no endereço www.cellrelay.com/cisco.htm.

[25] BOUNDEC, J.-Y. *The Assynchronous Transfer Mode: a tutorial. Computer Networks and OSDN Systems*. vol 27, n.1. Outubro 1992.

[26] *Sites* da *Internet*:

a. http://www.bellfone.com.br/ (Bellfone Telecomunicações Ltda.)
b. http://www.wirelessbrasil.org/
c. http://www.10emtudo.com.br/
d. http://www.atmforum.org/ (The ATM Fórum)
e. http://www.teleco.com.br/
f. http://www.itu.int/ (International Telecommunication Union)
g. http://www.pt.com/ (SS7 Tutorial by Performance Technologies)
h. http://www.elsevier.com/ (Elsevier: scientific, technical and health information)
i. http://www.nortel.com (Nortel Networks)
j. http://www.lucent.com (Lucent Technologies)
k. http://www.convergdigest.com/ (Converge! Network Digest)
l. http://www.mplsf.com/ (Fórum MPLS)
m. http://www.lucalm.hpg.ig.com.br/
n. http://www.cisco.com/